

EV333999509US

## NOVEL HUMAN GENES AND GENE EXPRESSION PRODUCTS

## CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/142,311, filed July 2, 1999, entitled "Novel Human Genes and Gene Expression Products" and also  
5 U.S. Provisional Application No. 60/142,310, filed July 2, 1999, entitled "Novel Human Genes and Gene Expression Products".

## FIELD OF THE INVENTION

The present invention relates to novel polynucleotides of human origin and the encoded gene products.

## 10 BACKGROUND OF THE INVENTION

Identification of novel polynucleotides, particularly those that encode an expressed gene product, is important in the advancement of drug discovery, diagnostic technologies, and the understanding of the progression and nature of complex diseases such as cancer. Identification of genes expressed in different cell types isolated from sources that differ in  
15 disease state or stage, developmental stage, exposure to various environmental factors, the tissue of origin, the species from which the tissue was isolated, and the like is key to identifying the genetic factors that are responsible for the phenotypes associated with these various differences.

This invention provides novel human polynucleotides, the polypeptides encoded by  
20 these polynucleotides, and the genes and proteins corresponding to these novel polynucleotides.

## SUMMARY OF THE INVENTION

This invention relates to novel human polynucleotides and variants thereof, their encoded polypeptides and variants thereof, to genes corresponding to these polynucleotides  
25 and to proteins expressed by the genes. The invention also relates to diagnostics and therapeutics comprising such novel human polynucleotides, their corresponding genes or gene products, including probes, antisense nucleotides, and antibodies. The polynucleotides of the invention correspond to a polynucleotide comprising the sequence information of at least one of SEQ ID NOs: 1-3351.

Various aspects and embodiments of the invention will be readily apparent to the ordinarily skilled artisan upon reading the description provided herein.

#### DETAILED DESCRIPTION OF THE INVENTION

The invention relates to polynucleotides comprising the disclosed nucleotide  
 5 sequences, to full length cDNA, mRNA genomic sequences, and genes corresponding to these sequences and degenerate variants thereof, and to polypeptides encoded by the polynucleotides of the invention and polypeptide variants.

Polypeptide variants differ from wild type protein in having one or more amino acid  
 10 substitutions that either enhance, add, or diminish a biological activity of the wild type protein.

Six of the polypeptides disclosed herein encode new members of the MKK kinase family; the coding region is found within the nucleotide region in parentheses: SEQ ID NO:29 (nucleotides 295-421); SEQ ID NO:31 (298-397); SEQ ID NO:196 (37-322); SEQ ID NO:3175 (nucleotides 14-164); SEQ ID NO:3190 (229-390); and SEQ ID NO:3281  
 15 (15-182). Twenty-four of the polypeptides encode new members of the family of transcription factor proteins having a basic region plus leucine zipper: SEQ ID NO:410 (42-191); SEQ ID NO:552 (116-288); SEQ ID NO:768 (116-288); SEQ ID NO:822 (108-262); SEQ ID NO:836 (158-353); SEQ ID NO:1288 (73-234); SEQ ID NO:1365 (69-257); SEQ ID NO:1540 (289-471); SEQ ID NO:1549 (200-391); SEQ ID NO:1556 (163-354);  
 20 SEQ ID NO:1557 (207-398); SEQ ID NO:1563 (107-298); SEQ ID NO:1622 (180-365); SEQ ID NO:1630 (100-291); SEQ ID NO:1704 (184-372); SEQ ID NO:1808 (36-161); SEQ ID NO:1454 (49-209); SEQ ID NO:2363 (48-211); SEQ ID NO:2424 (43-194); SEQ ID NO:3147 (190-369); SEQ ID NO:3152 (129-320); SEQ ID NO:3158 (167-334); and SEQ ID NO:3208 (34-256).

25 SEQ ID NOs:186 (175-395); 2591 (60-165); 3307 (43-321); and 3339 (94-342) encode polypeptides having an SH2 domain, and SEQ ID NOs:234 (23-121), 1832 (18-173), and 1835 (57-206) encode polypeptides having an SH3 domain. Nine polypeptides encode new members of the family of proteins having Ank repeat regions: SEQ ID NO:187 (358-432); SEQ ID NO:1268 (238-315); SEQ ID NO:1804 (301-378); SEQ ID NO:1819  
 30 (278-355); SEQ ID NO:1839 (224-307); SEQ ID NO:1830 (184-267); SEQ ID NO:2562 (18-101); SEQ ID NO:3015 (131-214); and SEQ ID NO:3267 (97-180).

The following eleven polynucleotides encode polypeptides having a C2H2 type zinc finger: SEQ ID NOs:308 (110-172); 807 (339-392); 1324 (294-356); 1503 (154-216);

1527 (156-212); 1674 (196-258); 1779 (64-126); 1801 (295-351); 3081 (190-252); 3193 (293-355); and 3306 (161-223). Eight polynucleotides encode polypeptides of the family of ATPases: SEQ ID NOs:431 (71-428); 639 (157-561); 2135 (2-401); 2684 (9-461); 2859 (100-320); 3178 (45-386); 3197 (281-343) and 3266 (8-139). Polypeptides having a  
 5 fibronectin type III domain are encoded by SEQ ID NO:746 (209-427) and 1192 (186-416). Polypeptides having an EF-hand domain are encoded by SEQ ID NO:820 (341-406); 1755 (281-367) and 3285(16-102). Six polypeptides of the protein kinase family are encoded by SEQ ID NOs:1157 (41-444); 1478 (54-437), 1496 (241-520); 2286 (12-182); 2969 (5-387); and 3190 (118-390).

10 LIM domain-containing polypeptides are encoded by SEQ ID NO:1269 (79-240); 1309 (248-404); 1360 (222-377); and 1386 (243-398). Two polypeptides of the family having a C2 domain (protein kinase C-like) are encoded by SEQ ID NO:1325 (1-234) and 2282(183-353). Polypeptides having a WD domain, G-beta repeat motif are encoded by SEQ ID NOs:1336 (66-164); 1380 (42-140); 1711 (263-361); 1762 (236-334); 1909 (160-  
 15 258); 2218 (127-225); 3047 (191-292); 3108 (275-367) and 3292 (208-300).

SEQ ID NO:1410 (222-350) encodes a member of the trypsin family. SEQ ID NOs:1417 (8-354); 2281 (20-387) and 2310 (20-371) encode members of the protein tyrosine phosphatase family. SEQ ID NOs:1464 (4-180) and 1514 (2-252) encode members of the family having an RNA recognition motif (also known as RRM, RBD, or  
 20 RNP domain). SEQ ID NOs:1496 (241-520) and 3297(7-153) encode helicases having a conserved C-terminal domain. SEQ ID NO:1538 (9-635) encodes a member of the wnt family of developmental signaling proteins.

Three polynucleotides encode polypeptides having a homeobox domain: SEQ ID NOs:1676 (9-86); 1820 (123-299); and 1821 (127-303). A novel thioredoxin is encoded by  
 25 SEQ ID NO:1677 (316-369). Two novel members of the ras family are encoded by SEQ ID NO:1688(109-410) and 3258(138-394). A novel polypeptide having a phosphatidylinositol-specific phospholipase C Y-domain is encoded by SEQ ID NO:1707 (92-439). A novel serine carboxypeptidase is encoded by SEQ ID NO:1744 (238-433). A novel polypeptide having N-terminal homology in the Ets domain is encoded by SEQ ID  
 30 NO:1811 (184-315). A novel polypeptide having a bromodomain is encoded by SEQ ID NO:1814 (127-294). A novel polypeptide having a double-stranded RNA binding motif is encoded by SEQ ID NO:1818 (9-146). A novel polypeptide having a G-protein alpha subunit is encoded by SEQ ID NO:1846 (12-398).

SEQ ID NOs:1911 (35-151) and 1980 (60-197) encode polypeptides having a  
 35 C3HC4 type zinc finger domain (RING finger). SEQ ID NO:2065 (253-306) encodes a

polypeptide having a CCHC zinc finger domain. SEQ ID NO:2216 (90-179) encodes a polypeptide having a WW/rsp5/WWP domain. SEQ ID NO:2428 (25-350) encodes a polypeptide member of the dual specificity phosphatase family, having a catalytic domain.

SEQ ID NOs:2577 (0-311); 3183 (14-215); and 3195 (0-215) encode members of the 4 transmembrane segment integral membrane protein family. SEQ ID NOs:2826 (116-400) and 2871 (198-392) encode polypeptides of the DEAD and DEAH box helicase family. SEQ ID NO:2944 (18-281) encodes a polypeptide having a calpain large subunit, domain III.

SEQ ID NO:3274 (11-187) encodes a eukaryotic transcription factor with a fork head domain. SEQ ID NO:3345 (65-271) encodes a polypeptide having a PDZ domain, and SEQ ID NO:3351 (124-270) encodes a polypeptide in the family of phorbol esters/glycerol binding proteins.

Described below are polynucleotide compositions encompassed by the invention, methods for obtaining cDNA or genomic DNA encoding a full-length gene product, expression of these polynucleotides and genes, identification of structural motifs of the polynucleotides and genes, identification of the function of a gene product encoded by a gene corresponding to a polynucleotide of the invention, use of the provided polynucleotides as probes and in mapping and in tissue profiling, use of the corresponding polypeptides and other gene products to raise antibodies, and use of the polynucleotides and their encoded gene products for therapeutic and diagnostic purposes.

#### Polynucleotide Compositions

The scope of the invention with respect to polynucleotide compositions includes, but is not necessarily limited to, polynucleotides having a sequence set forth in any one of SEQ ID NOs:1-3351; polynucleotides obtained from the biological materials described herein or other biological sources (particularly human sources) by hybridization under stringent conditions (particularly conditions of high stringency); genes corresponding to the provided polynucleotides; variants of the provided polynucleotides and their corresponding genes, particularly those variants that retain a biological activity of the encoded gene product (*e.g.*, a biological activity ascribed to a gene product corresponding to the provided polynucleotides as a result of the assignment of the gene product to a protein family(ies) and/or identification of a functional domain present in the gene product). Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here.



“Polynucleotide” and “nucleic acid” as used herein with reference to nucleic acids of the composition is not intended to be limiting as to the length or structure of the nucleic acid unless specifically indicated.

The invention features polynucleotides that are expressed in human tissue,  
 5 specifically human colon, breast, and/or lung tissue. Novel nucleic acid compositions of the invention comprise a sequence set forth in any one of SEQ ID NOs:1-3351 or an identifying sequence thereof. An “identifying sequence” is a contiguous sequence of residues at least about 10 nt to about 20 nt in length, usually at least about 50 nt to about 100 nt in length, that uniquely identifies a polynucleotide sequence, *e.g.*, exhibits less than  
 10 90%, usually less than about 80% to about 85% sequence identity to any contiguous nucleotide sequence of more than about 20 nt. Thus, the subject novel nucleic acid compositions include full length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of SEQ ID NOs:1-3351.

The polynucleotides of the invention also include polynucleotides having sequence  
 15 similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1XSSC (9 mM saline/0.9 mM sodium citrate).  
 20 Hybridization methods and conditions are well known in the art, see, *e.g.*, U.S. Patent No. 5,707,829. Nucleic acids that are substantially identical to the provided polynucleotide sequences, *e.g.*, allelic variants, genetically altered versions of the gene, *etc.*, bind to the provided polynucleotide sequences (SEQ ID NOs:1-3351) under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can  
 25 isolate homologous or related genes. The source of homologous genes can be any species, *e.g.*, primate species, particularly human; rodents, such as rats and mice; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.*

Preferably, hybridization is performed using at least 15 contiguous nucleotides (nt) of at least one of SEQ ID NOs:1-3351. That is, when at least 15 contiguous nt of one of the  
 30 disclosed SEQ ID NOs. is used as a probe, the probe will preferentially hybridize with a nucleic acid comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids that uniquely hybridize to the selected probe. Probes from more than one SEQ ID NO. can hybridize with the same nucleic acid if the cDNA from which they were derived corresponds to one mRNA. Probes of more than 15 nt can be

used, *e.g.*, probes of from about 18 nt to about 100 nt, but 15 nt represents sufficient sequence for unique identification.

The polynucleotides of the invention also include naturally occurring variants of the nucleotide sequences (*e.g.*, degenerate variants, allelic variants). Variants of the polynucleotides of the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the polynucleotides of the invention can be identified where the allelic variant exhibits at most about 25-30% base pair (bp) mismatches relative to the selected polynucleotide probe. In general, allelic variants contain 15-25% bp mismatches, and can contain as little as even 5-15%, or 2-5%, or 1-2% bp mismatches, as well as a single bp mismatch.

The invention also encompasses homologs corresponding to the polynucleotides of SEQ ID NOs:1-3351, where the source of homologous genes can be any mammalian species, *e.g.*, primate species, particularly human; rodents, such as rats; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.* Between mammalian species, *e.g.*, human and mouse, homologs generally have substantial sequence similarity, *e.g.*, at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about 18 contiguous nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul et al., *J. Mol. Biol.* (1990) 215:403-10.

In general, variants of the invention have a sequence identity greater than at least about 65%, preferably at least about 75%, more preferably at least about 85%, and can be greater than at least about 90%, 91%, 92%, 93%, 94%, 95%, or 96%, most preferably 97%, 98% or 99%. For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm, using the following. Global DNA sequence identity must be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty, 12; and gap extension penalty, 1.

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein (*e.g.*, in diagnosis, as a unique identifier of a

differentially expressed gene of interest, *etc.*). The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide of the invention.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression.

The nucleic acid compositions of the subject invention can encode all or a part of the subject polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* Isolated polynucleotides and polynucleotide fragments of the invention comprise at least about 10, about 15, about 20, about 35, about 50, about 100, about 150 to about 200, about 250 to about 300, or about 350 contiguous nt selected from the polynucleotide sequences as shown in SEQ ID NOs:1-3351. The fragments also include those of lengths intermediate to the specifically mentioned lengths, such as 35, 36, 37, 38, 39, *etc.*; 150, 151, 152, 153, 154, *etc.* For the most part, fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and up to at least about 50 contiguous nt in length or more. In a preferred embodiment, the polynucleotide molecules comprise a contiguous sequence of at least 12 nt selected from the group consisting of the polynucleotides shown in SEQ ID NOs:1-3351.

Probes specific to the polynucleotides of the invention can be generated using the polynucleotide sequences disclosed in SEQ ID NOs:1-3351. The probes are preferably at least about a 12, 15, 16, 18, 20, 22, 24, or 25 nt fragment of a corresponding contiguous sequence of SEQ ID NOs:1-3351, and can be less than 2, 1, 0.5, 0.1, or 0.05 kb in length. The probes can be synthesized chemically or can be generated from longer polynucleotides

using restriction enzymes. The probes can be labeled, for example, with a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of a polynucleotide of one of SEQ ID NOs:1-3351. More preferably, probes are designed based on a contiguous sequence of one of the subject polynucleotides that remain  
 5 unmasked following application of a masking program for masking low complexity (*e.g.*, XBLAST) to the sequence., *i.e.*, one would select an unmasked region, as indicated by the polynucleotides outside the poly-n stretches of the masked sequence produced by the masking program.

The polynucleotides of the subject invention are isolated and obtained in substantial  
 10 purity, generally as other than an intact chromosome. Usually, the polynucleotides, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

15 The polynucleotides of the invention can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the polynucleotides can be regulated by their own or by other regulatory sequences known in the art. The polynucleotides of the invention can be introduced into suitable host cells  
 20 using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

25 The subject nucleic acid compositions can be used to, for example, produce polypeptides, as probes for the detection of mRNA of the invention in biological samples (*e.g.*, extracts of human cells) to generate additional copies of the polynucleotides, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for  
 30 example, determine the presence or absence of the polynucleotide sequences as shown in SEQ ID NOs:1-3351 or variants thereof in a sample. These and other uses are described in more detail below.

### Use of Polynucleotides to Obtain Full-Length cDNA, Gene, and Promoter Region

Full-length cDNA molecules comprising the disclosed polynucleotides are obtained as follows. A polynucleotide having a sequence of one of SEQ ID NOs:1-3351, or a portion thereof comprising at least 12, 15, 18, or 20 nt, is used as a hybridization probe to  
 5 detect hybridizing members of a cDNA library using probe design methods, cloning methods, and clone selection techniques such as those described in U.S. Patent No. 5,654,173. Libraries of cDNA are made from selected tissues, such as normal or tumor tissue, or from tissues of a mammal treated with, for example, a pharmaceutical agent. Preferably, the tissue is the same as the tissue from which the polynucleotides of the  
 10 invention were isolated, as both the polynucleotides described herein and the cDNA represent expressed genes. Most preferably, the cDNA library is made from the biological material described herein in the Examples. The choice of cell type for library construction can be made after the identity of the protein encoded by the gene corresponding to the polynucleotide of the invention is known. This will indicate which tissue and cell types are  
 15 likely to express the related gene, and thus represent a suitable source for the mRNA for generating the cDNA. As described in the Examples, cDNA of the invention was isolated from specific cell or tissue types, and such cells and tissues are preferable for obtaining related nucleic acids.

Techniques for producing and probing nucleic acid sequence libraries are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual, 2nd Ed.*,  
 20 (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. The cDNA can be prepared by using primers based on sequence from SEQ ID NOs:1-3351. In one embodiment, the cDNA library can be made from only poly-adenylated mRNA. Thus, poly-T primers can be used to prepare cDNA from the mRNA.

Members of the library that are larger than the provided polynucleotides, and preferably that encompass the complete coding sequence of the native message, are obtained. In order to confirm that the entire cDNA has been obtained, RNA protection experiments are performed as follows. Hybridization of a full-length cDNA to an mRNA will protect the RNA from RNase degradation. If the cDNA is not full length, then the  
 25 portions of the mRNA that are not hybridized will be subject to RNase degradation. This is assayed, as is known in the art, by changes in electrophoretic mobility on polyacrylamide gels, or by detection of released monoribonucleotides. Sambrook et al., *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. In order to obtain additional sequences 5' to the end of a partial cDNA, 5'  
 30

RACE (*PCR Protocols: A Guide to Methods and Applications*, (1990) Academic Press, Inc.) can be performed.

Genomic DNA is isolated using the provided polynucleotides in a manner similar to the isolation of full-length cDNAs. Briefly, the provided polynucleotides, or portions thereof, are used as probes to libraries of genomic DNA. Preferably, the library is obtained from the cell type that was used to generate the polynucleotides of the invention, but this is not essential. Most preferably, the genomic DNA is obtained from the biological material described herein in the Examples. Such libraries can be in vectors suitable for carrying large segments of a genome, such as P1 or YAC, as described in detail in Sambrook et al., 9.4-9.30. In addition, genomic sequences can be isolated from human BAC libraries, which are commercially available from Research Genetics, Inc., Huntsville, Alabama, USA, for example. In order to obtain additional 5' or 3' sequences, chromosome walking is performed, as described in Sambrook et al., such that adjacent and overlapping fragments of genomic DNA are isolated. These are mapped and pieced together, as is known in the art, using restriction digestion enzymes and DNA ligase.

Using the polynucleotide sequences of the invention, corresponding full-length genes can be isolated using both classical and PCR methods to construct and probe cDNA libraries. Using either method, Northern blots, preferably, are performed on a number of cell types to determine which cell lines express the gene of interest at the highest level. Classical methods of constructing cDNA libraries are taught in Sambrook et al., *supra*. With these methods, cDNA can be produced from mRNA and inserted into viral or expression vectors. Typically, libraries of mRNA comprising poly(A) tails can be produced with poly(T) primers. Similarly, cDNA libraries can be produced using the instant sequences as primers.

PCR methods are used to amplify the members of a cDNA library that comprise the desired insert. In this case, the desired insert will contain sequence from the full length cDNA that corresponds to the instant polynucleotides. Such PCR methods include gene trapping and RACE methods as described in Gruber et al., WO 95/04745 and Gruber et al., U.S. Patent No. 5,500,356. Kits are commercially available to perform gene trapping experiments from, for example, Life Technologies, Gaithersburg, Maryland, USA. In preferred embodiments of RACE, a common primer is designed to anneal to an arbitrary adaptor sequence ligated to cDNA ends (Apte and Siebert, *Biotechniques* (1993) 15:890-893; Edwards et al., *Nuc. Acids Res.* (1991) 19:5227-5232). When a single gene-specific RACE primer is paired with the common primer, preferential amplification of sequences

between the single gene specific primer and the common primer occurs. Commercial cDNA pools modified for use in RACE are available.

The promoter region of a gene generally is located 5' to the initiation site for RNA polymerase II. Hundreds of promoter regions contain the "TATA" box, a sequence such as TATTA or TATAA, which is sensitive to mutations. The promoter region can be obtained by performing 5' RACE using a primer from the coding region of the gene. Alternatively, the cDNA can be used as a probe for the genomic sequence, and the region 5' to the coding region is identified by "walking up." If the gene is highly expressed or differentially expressed, the promoter from the gene can be of use in a regulatory construct for a heterologous gene.

Once the full-length cDNA or gene is obtained, DNA encoding variants can be prepared by site-directed mutagenesis, described in detail in Sambrook et al., 15.3-15.63. The choice of codon or nucleotide to be replaced can be based on disclosure herein on optional changes in amino acids to achieve altered protein structure and/or function.

As an alternative method to obtaining DNA or RNA from a biological material, nucleic acid comprising nucleotides having the sequence of one or more polynucleotides of the invention can be synthesized. Thus, the invention encompasses nucleic acid molecules ranging in length from 15 nt (corresponding to at least 15 contiguous nt of one of SEQ ID NOs:1-3351) up to a maximum length suitable for one or more biological manipulations, including replication and expression, of the nucleic acid molecule. The invention includes but is not limited to (a) nucleic acid having the size of a full gene, and comprising at least one of SEQ ID NOs:1-3351; (b) the nucleic acid of (a) also comprising at least one additional polynucleotide or gene, operably linked to permit expression of a fusion protein; (c) an expression vector comprising (a) or (b); (d) a plasmid comprising (a) or (b); and (e) a recombinant viral particle comprising (a) or (b). Once provided with the polynucleotides disclosed herein, construction or preparation of (a) - (e) are well within the skill in the art.

The sequence of a nucleic acid comprising at least 15 contiguous nt of at least any one of SEQ ID NOs:1-3351, preferably the entire sequence of at least any one of SEQ ID NOs:1-3351, is not limited and can be any sequence of A, T, G, and/or C (for DNA) and A, U, G, and/or C (for RNA) or modified bases thereof, including inosine and pseudouridine. The choice of sequence will depend on the desired function and can be dictated by coding regions desired, the intron-like regions desired, and the regulatory regions desired. Where the entire sequence of any one of SEQ ID NOs:1-3351 is within the nucleic acid, the nucleic acid obtained is referred to herein as a polynucleotide comprising the sequence of any one of SEQ ID NOs:1-3351.

### Expression of Polypeptide Encoded by Full-Length cDNA or Full-Length Gene

The provided polynucleotides (*e.g.*, a polynucleotide having a sequence of one of SEQ ID NOs:1-3351), the corresponding cDNA, or the full-length gene is used to express a partial or complete gene product. Constructs of polynucleotides having sequences of SEQ ID NOs:1-3351 can be generated synthetically. Alternatively, single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, *e.g.*, Stemmer et al., *Gene (Amsterdam)* (1995) 164(1):49-53. In this method, assembly PCR (the synthesis of long DNA sequences from large numbers of oligodeoxyribonucleotides (oligos)) is described. The method is derived from DNA shuffling (Stemmer, *Nature* (1994) 370:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process.

Appropriate polynucleotide constructs are purified using standard recombinant DNA techniques as described in, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, and under current regulations described in United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research. The gene product encoded by a polynucleotide of the invention is expressed in any expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems. Vectors, host cells and methods for obtaining expression in same are well known in the art. Suitable vectors and host cells are described in U.S. Patent No. 5,654,173.

Polynucleotide molecules comprising a polynucleotide sequence provided herein are generally propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole animal or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially. Methods for preparation of vectors comprising a desired sequence are well known in the art.

The polynucleotides set forth in SEQ ID NOs:1-3351 or their corresponding full-length polynucleotides are linked to regulatory sequences as appropriate to obtain the desired expression properties. These can include promoters (attached either at the 5' end of the sense strand or at the 3' end of the antisense strand), enhancers, terminators, operators, repressors, and inducers. The promoters can be regulated or constitutive. In some



situations it may be desirable to use conditionally active promoters, such as tissue-specific or developmental stage-specific promoters. These are linked to the desired nucleotide sequence using the techniques described above for linkage to vectors. Any techniques known in the art can be used.

5           When any appropriate host cells or organisms are used to replicate and/or express the polynucleotides or nucleic acids of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

10           Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the cell to which the gene is native. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence as disclosed in U.S. Patent No. 5,641,670.

#### Identification of Functional and Structural Motifs of Novel Genes

15           Translations of the nucleotide sequence of the provided polynucleotides, cDNAs or full genes can be aligned with individual known sequences. Similarity with individual sequences can be used to determine the activity of the polypeptides encoded by the polynucleotides of the invention. Also, sequences exhibiting similarity with more than one individual sequence can exhibit activities that are characteristic of either or both individual sequences.

20           The full length sequences and fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence corresponding to provided polynucleotides. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences corresponding to the provided polynucleotides.

25           Typically, a selected polynucleotide is translated in all six frames to determine the best alignment with the individual sequences. The sequences disclosed herein in the Sequence Listing are in a 5' to 3' orientation and translation in three frames can be sufficient. These amino acid sequences are referred to, generally, as query sequences, which will be aligned with the individual sequences. Databases with individual sequences are described in "Computer Methods for Macromolecular Sequence Analysis" *Methods in Enzymology* (1996) 266, Doolittle, Academic Press, Inc., a division of Harcourt Brace &

Co., San Diego, California, USA. Databases include Genbank, EMBL, and DNA Database of Japan (DDBJ).

Query and individual sequences can be aligned using the methods and computer programs described above, and include BLAST, available over the world wide web at  
 5 <http://www.ncbi.nlm.nih.gov/BLAST>. Another alignment algorithm is Fasta, available in the Genetics Computing Group (GCG) package, Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc. Other techniques for alignment are described in Doolittle, *supra*. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm  
 10 that permits gaps in sequence alignments. See *Meth. Mol. Biol.* (1997) 70: 173-187. Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to identify sequences that are  
 15 distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Amino acid sequences encoded by the provided polynucleotides can be used to search both protein and DNA databases.

High Similarity. In general, in alignment results considered to be of high similarity, the percent of the alignment region length is typically at least about 55% of total length  
 20 query sequence; more typically, at least about 58%; even more typically; at least about 60% of the total residue length of the query sequence. Usually, percent length of the alignment region can be as much as about 62%; more usually, as much as about 64%; even more usually, as much as about 66%. Further, for high similarity, the region of alignment, typically, exhibits at least about 75% of sequence identity; more typically, at least about  
 25 78%; even more typically; at least about 80% sequence identity. Usually, percent sequence identity can be as much as about 82%; more usually, as much as about 84%; even more usually, as much as about 86%.

The p value is used in conjunction with these methods. If high similarity is found, the query sequence is considered to have high similarity with a profile sequence when the p  
 30 value is less than or equal to about  $10^{-2}$ ; more usually; less than or equal to about  $10^{-3}$ ; even more usually; less than or equal to about  $10^{-4}$ . More typically, the p value is no more than about  $10^{-5}$ ; more typically; no more than or equal to about  $10^{-10}$ ; even more typically; no more than or equal to about  $10^{-15}$  for the query sequence to be considered high similarity.

Similarity Determined by Sequence Identity Alone. Sequence identity alone can be  
 35 used to determine similarity of a query sequence to an individual sequence and can indicate

the activity of the sequence. Such an alignment, preferably, permits gaps to align sequences. Typically, the query sequence is related to the profile sequence if the sequence identity over the entire query sequence is at least about 15%; more typically, at least about 20%; even more typically, at least about 25%; even more typically, at least about 50%.

- 5 Sequence identity alone as a measure of similarity is most useful when the query sequence is usually, at least 80 residues in length; more usually, 90 residues; even more usually, at least 95 amino acid residues in length. More typically, similarity can be concluded based on sequence identity alone when the query sequence is preferably 100 residues in length; more preferably, 120 residues in length; even more preferably, 150 amino acid residues in  
10 length.

- Alignments with Profile and Multiple Aligned Sequences. Translations of the provided polynucleotides can be aligned with amino acid profiles that define either protein families or common motifs. Also, translations of the provided polynucleotides can be aligned to multiple sequence alignments (MSA) comprising the polypeptide sequences of  
15 members of protein families or motifs. Similarity or identity with profile sequences or MSAs can be used to determine the activity of the gene products (*e.g.*, polypeptides) encoded by the provided polynucleotides or corresponding cDNA or genes. For example, sequences that show an identity or similarity with a chemokine profile or MSA can exhibit chemokine activities.

- 20 Profiles can be designed manually by (1) creating an MSA, which is an alignment of the amino acid sequence of members that belong to the family and (2) constructing a statistical representation of the alignment. Such methods are described, for example, in Birney et al., *Nucl. Acid Res.* (1996) 24(14): 2730-2739. MSAs of some protein families and motifs are publicly available. MSAs are described also in Sonnhammer et al., *Proteins*  
25 (1997) 28: 405-420. A brief description of MSAs is reported in Pascarella et al., *Prot. Eng.* (1996) 9(3):249-251. Techniques for building profiles from MSAs are described in Sonnhammer et al., *supra*; Birney et al., *supra*; and "Computer Methods for Macromolecular Sequence Analysis," *Methods in Enzymology* (1996) 266, Doolittle, Academic Press, Inc., San Diego, California, USA.

- 30 Similarity between a query sequence and a protein family or motif can be determined by (a) comparing the query sequence against the profile and/or (b) aligning the query sequence with the members of the family or motif. Typically, a program such as Searchwise is used to compare the query sequence to the statistical representation of the multiple alignment, also known as a profile (see Birney et al., *supra*). Other techniques to

compare the sequence and profile are described in Sonnhammer et al., *supra* and Doolittle, *supra*.

Next, methods described by Feng et al., *J. Mol. Evol.* (1987) 25:351 and Higgins et al., *CABIOS* (1989) 5:151 can be used align the query sequence with the members of a family or motif, also known as a MSA. Sequence alignments can be generated using any of a variety of software tools. Examples include PileUp, which creates a multiple sequence alignment, and is described in Feng et al., *J. Mol. Evol.* (1987) 25:351. Another method, GAP, uses the alignment method of Needleman et al., *J. Mol. Biol.* (1970) 48:443. GAP is best suited for global alignment of sequences. A third method, BestFit, functions by inserting gaps to maximize the number of matches using the local homology algorithm of Smith et al., *Adv. Appl. Math.* (1981) 2:482. In general, the following factors are used to determine if a similarity between a query sequence and a profile or MSA exists: (1) number of conserved residues found in the query sequence, (2) percentage of conserved residues found in the query sequence, (3) number of frameshifts, and (4) spacing between conserved residues.

Some alignment programs that both translate and align sequences can make any number of frameshifts when translating the nucleotide sequence to produce the best alignment. The fewer frameshifts needed to produce an alignment, the stronger the similarity or identity between the query and profile or MSAs. For example, a weak similarity resulting from no frameshifts can be a better indication of activity or structure of a query sequence, than a strong similarity resulting from two frameshifts. Preferably, three or fewer frameshifts are found in an alignment; more preferably two or fewer frameshifts; even more preferably, one or fewer frameshifts; even more preferably, no frameshifts are found in an alignment of query and profile or MSAs.

Conserved residues are those amino acids found at a particular position in all or some of the family or motif members. Alternatively, a position is considered conserved if only a certain class of amino acids is found in a particular position in all or some of the family members. For example, the N-terminal position can contain a positively charged amino acid, such as lysine, arginine, or histidine.

Typically, a residue of a polypeptide is conserved when a class of amino acids or a single amino acid is found at a particular position in at least about 40% of all class members; more typically, at least about 50%; even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

A residue is considered conserved when three unrelated amino acids are found at a particular position in the some or all of the members; more usually, two unrelated amino acids. These residues are conserved when the unrelated amino acids are found at particular positions in at least about 40% of all class member; more typically, at least about 50%;  
 5 even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

10 A query sequence has similarity to a profile or MSA when the query sequence comprises at least about 25% of the conserved residues of the profile or MSA; more usually, at least about 30%; even more usually; at least about 40%. Typically, the query sequence has a stronger similarity to a profile sequence or MSA when the query sequence comprises at least about 45% of the conserved residues of the profile or MSA; more typically, at least about 50%; even more typically; at least about 55%.

#### 15 Identification of Secreted and Membrane-Bound Polypeptides

Both secreted and membrane-bound polypeptides of the present invention are of particular interest. For example, levels of secreted polypeptides can be assayed in body fluids that are convenient, such as blood, plasma, serum, and other body fluids such as urine, prostatic fluid and semen. Membrane-bound polypeptides are useful for constructing  
 20 vaccine antigens or inducing an immune response. Such antigens would comprise all or part of the extracellular region of the membrane-bound polypeptides. Because both secreted and membrane-bound polypeptides comprise a fragment of contiguous hydrophobic amino acids, hydrophobicity predicting algorithms can be used to identify such polypeptides.

25 A signal sequence is usually encoded by both secreted and membrane-bound polypeptide genes to direct a polypeptide to the surface of the cell. The signal sequence usually comprises a stretch of hydrophobic residues. Such signal sequences can fold into helical structures. Membrane-bound polypeptides typically comprise at least one transmembrane region that possesses a stretch of hydrophobic amino acids that can  
 30 transverse the membrane. Some transmembrane regions also exhibit a helical structure. Hydrophobic fragments within a polypeptide can be identified by using computer algorithms. Such algorithms include Hopp & Woods, *Proc. Natl. Acad. Sci. USA* (1981)

78:3824-3828; Kyte & Doolittle, *J. Mol. Biol.* (1982) 157: 105-132; and RAOAR algorithm, Degli Esposti et al., *Eur. J. Biochem.* (1990) 190: 207-219.

Another method of identifying secreted and membrane-bound polypeptides is to translate the polynucleotides of the invention in all six frames and determine if at least 8 contiguous hydrophobic amino acids are present. Those translated polypeptides with at least 8; more typically, 10; even more typically, 12 contiguous hydrophobic amino acids are considered to be either a putative secreted or membrane bound polypeptide. Hydrophobic amino acids include alanine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine

#### 10 Identification of the Function of an Expression Product of a Full-Length Gene

Ribozymes, antisense constructs, and dominant negative mutants can be used to determine function of the expression product of a gene corresponding to a polynucleotide provided herein. The phosphoramidite method of oligonucleotide synthesis can be used to construct antisense molecules and ribozymes. See Beaucage et al., *Tet. Lett.* (1981) 22:1859 and U.S. Patent No. 4,668,777. Automated devices for synthesis are available to create oligonucleotides using this chemistry. Examples of such devices include Biosearch 8600, Models 392 and 394 by Applied Biosystems, a division of Perkin-Elmer Corp., Foster City, California, USA; and Expedite by Perceptive Biosystems, Framingham, Massachusetts, USA. Synthetic RNA, phosphate analog oligonucleotides, and chemically derivatized oligonucleotides can also be produced, and can be covalently attached to other molecules. RNA oligonucleotides can be synthesized, for example, using RNA phosphoramidites. This method can be performed on an automated synthesizer, such as Applied Biosystems, Models 392 and 394, Foster City, California, USA.

Oligonucleotides of up to 200 nt can be synthesized, more typically, 100 nt, more typically 50 nt; even more typically 30 to 40 nt. These synthetic fragments can be annealed and ligated together to construct larger fragments. See, for example, Sambrook et al., *supra*. Trans-cleaving catalytic RNAs (ribozymes) are RNA molecules possessing endoribonuclease activity. Ribozymes are specifically designed for a particular target, and the target message must contain a specific nucleotide sequence. They are engineered to cleave any RNA species site-specifically in the background of cellular RNA. The cleavage event renders the mRNA unstable and prevents protein expression. Importantly, ribozymes can be used to inhibit expression of a gene of unknown function for the purpose of determining its function in an *in vitro* or *in vivo* context, by detecting the phenotypic effect.

Antisense nucleic acids are designed to specifically bind to RNA, resulting in the formation of RNA-DNA or RNA-RNA hybrids, with an arrest of DNA replication, reverse transcription or messenger RNA translation. Antisense polynucleotides based on a selected polynucleotide sequence can interfere with expression of the corresponding gene.

5 Antisense polynucleotides are typically generated within the cell by expression from antisense constructs that contain the antisense strand as the transcribed strand. Antisense polynucleotides based on the disclosed polynucleotides will bind and/or interfere with the translation of mRNA comprising a sequence complementary to the antisense polynucleotide. The expression products of control cells and cells treated with the  
10 antisense construct are compared to detect the protein product of the gene corresponding to the polynucleotide upon which the antisense construct is based. The protein is isolated and identified using routine biochemical methods.

Given the extensive background literature and clinical experience in antisense therapy, one skilled in the art can use selected polynucleotides of the invention as  
15 additional potential therapeutics. The choice of polynucleotide can be narrowed by first testing them for binding to "hot spot" regions of the genome of cancerous cells. If a polynucleotide is identified as binding to a "hot spot," testing the polynucleotide as an antisense compound in the corresponding cancer cells is warranted.

Dominant negative mutations also are readily generated for corresponding proteins  
20 that are active as homomultimers. A mutant polypeptide will interact with wild-type polypeptides (made from the other allele) and form a non-functional multimer. Thus, a mutation is in a substrate-binding domain, a catalytic domain, or a cellular localization domain. Preferably, the mutant polypeptide will be overproduced. Point mutations are made that have such an effect. In addition, fusion of different polypeptides of various  
25 lengths to the terminus of a protein can yield dominant negative mutants. General strategies are available for making dominant negative mutants (see, *e.g.*, Herskowitz, *Nature* (1987) 329:219). Such techniques can be used to create loss of function mutations, which are useful for determining protein function.

#### Polypeptides and Variants Thereof

30 The polypeptides of the invention include those encoded by the disclosed polynucleotides, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed polynucleotides. Thus, the invention

includes within its scope a polypeptide encoded by a polynucleotide having the sequence of any one of SEQ ID NOs:1-3351 or a variant thereof.

In general, the term “polypeptide” as used herein refers to both the full length polypeptide encoded by the recited polynucleotide, the polypeptide encoded by the gene represented by the recited polynucleotide, as well as portions or fragments thereof. “Polypeptides” also includes variants of the naturally occurring proteins, where such variants are homologous or substantially similar to the naturally occurring protein, and can be of an origin of the same or different species as the naturally occurring protein (*e.g.*, human, murine, or some other species that naturally expresses the recited polypeptide, usually a mammalian species). In general, variant polypeptides have a sequence that has at least about 80%, usually at least about 90%, and more usually at least about 98% sequence identity with a differentially expressed polypeptide of the invention, as measured by BLAST using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, *i.e.*, the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

The invention also encompasses homologs of the disclosed polypeptides (or fragments thereof) where the homologs are isolated from other species, *i.e.*, other animal or plant species, where such homologs, usually mammalian species, *e.g.*, rodents, such as mice, rats; domestic animals, *e.g.*, horse, cow, dog, cat; and humans. By “homolog” is meant a polypeptide having at least about 35%, usually at least about 40% and more usually at least about 60% amino acid sequence identity to a particular differentially expressed protein as identified above, where sequence identity is determined using the BLAST algorithm, with the parameters described above.

In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment, *e.g.*, are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition that is enriched for the protein as compared to a control. As such, purified polypeptide is provided, where by purified is meant that the protein is present in a composition that is substantially free of non-differentially expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of non-differentially expressed polypeptides.

Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a



phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/hydrophilicity, and/or steric bulk of the amino acid substituted. Variants can be designed so as to retain biological activity of a particular region of the protein (e.g., a functional domain and/or, where the polypeptide is a member of a protein family, a region associated with a consensus sequence). Selection of amino acid alterations for production of variants can be based upon the accessibility (interior vs. exterior) of the amino acid (see, e.g., Go et al., *Int. J. Peptide Protein Res.* (1980) 15:211), the thermostability of the variant polypeptide (see, e.g., Querol et al., *Prot. Eng.* (1996) 9:265), desired glycosylation sites (see, e.g., Olsen and Thomsen, *J. Gen. Microbiol.* (1991) 137:579), desired disulfide bridges (see, e.g., Clarke et al., *Biochemistry* (1993) 32:4322; and Wakarchuk et al., *Protein Eng.* (1994) 7:1379), desired metal binding sites (see, e.g., Toma et al., *Biochemistry* (1991) 30:97, and Haezerebrouck et al., *Protein Eng.* (1993) 6:643), and desired substitutions with in proline loops (see, e.g., Masul et al., *Appl. Env. Microbiol.* (1994) 60:3579). Cysteine-depleted muteins can be produced as disclosed in U.S. Patent No. 4,959,314.

Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 aa to at least about 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a polynucleotide having a sequence of any SEQ ID NOs:1-3351, or a homolog thereof. The protein variants described herein are encoded by polynucleotides that are within the scope of the invention. The genetic code can be used to select the appropriate codons to construct the corresponding variants.

#### Computer-Related Embodiments

In general, a library of polynucleotides is a collection of sequence information, which information is provided in either biochemical form (e.g., as a collection of polynucleotide molecules), or in electronic form (e.g., as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The sequence information of the polynucleotides can be used in a

variety of ways, *e.g.*, as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (*e.g.*, cell type markers), and/or as markers of a given disease or disease state. In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level  
 5 relative to a normal cell (*e.g.*, a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either overexpressed or underexpressed in a breast ductal cell affected by cancer relative to a normal (*i.e.*, substantially disease-free) breast cell.

10 The nucleotide sequence information of the library can be embodied in any suitable form, *e.g.*, electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file (or, in biochemical form, a collection of nucleic acid molecules) that contains the representative nucleotide sequences of genes that are differentially expressed (*e.g.*, overexpressed or underexpressed)  
 15 as between, for example, i) a cancerous cell and a normal cell; ii) a cancerous cell and a dysplastic cell; iii) a cancerous cell and a cell affected by a disease or condition other than cancer; iv) a metastatic cancerous cell and a normal cell and/or non-metastatic cancerous cell; v) a malignant cancerous cell and a non-malignant cancerous cell (or a normal cell) and/or vi) a dysplastic cell relative to a normal cell. Other combinations and comparisons  
 20 of cells affected by various diseases or stages of disease will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acids that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

25 The polynucleotide libraries of the subject invention generally comprise sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence of any of SEQ ID NOs:1-3351. By plurality is meant at least 2, usually at least 3 and can include up to all of SEQ ID NOs:1-3351. The length and number of polynucleotides in the library will vary with the nature of the library, *e.g.*, if the  
 30 library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, *etc.*

Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention.  
 35 Such a manufacture provides the genome sequence or a subset thereof in a form that can be

examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, *e.g.*, the nucleic acid sequences of any of the polynucleotides of SEQ ID NOs:1-3351, can be recorded on computer readable media, *e.g.*, any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, *e.g.*, word processing text file, database format, *etc.* In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (*e.g.*, searchable files, executable files, *etc.*, including, but not limited to, for example, search program software, *etc.*).

By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the BLAST (Altschul et al., *supra.*) and BLAZE (Brutlag et al. *Comp. Chem.* (1993) 17:203) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

“Search means” refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, *e.g.*, MacPattern (EMBL), BLASTN and BLASTX (NCBI). A “target sequence” can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nt. A variety of comparing means can be used to accomplish comparison of sequence information from a sample (*e.g.*, to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art.

A “target structural motif,” or “target motif,” refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile.

As discussed above, the “library” of the invention also encompasses biochemical libraries of the polynucleotides of SEQ ID NOs:1-3351, *e.g.*, collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, *e.g.*, a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (*i.e.*, an array) and the like. Of particular interest are nucleic acid arrays in which one or more of SEQ ID NOs:1-3351 is represented on the array. By array is meant an article of manufacture that has at least a substrate with at least two distinct

nucleic acid targets on one of its surfaces, where the number of distinct nucleic acids can be considerably higher, typically being at least 10 nt, usually at least 20 nt and often at least 25 nt. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the where the polypeptides of the library will represent at least a portion of the polypeptides encoded by SEQ ID NOs:1-3351.

#### 10 Use of Polynucleotide Probes in Mapping, and in Tissue Profiling

Polynucleotide probes, generally comprising at least 12 contiguous nt of a polynucleotide as shown in the Sequence Listing, are used for a variety of purposes, such as chromosome mapping of the polynucleotide and detection of transcription levels. Additional disclosure about preferred regions of the disclosed polynucleotide sequences is found in the Examples. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 5-, 10-, or 20-fold higher than the background hybridization provided with other unrelated sequences.

Detection of Expression Levels. Nucleotide probes are used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots, mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization is quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for *in situ* hybridization to cells to detect expression. Probes can also be used *in vivo* for diagnostic detection of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores, fluors, and enzymes. Other examples of nucleotide hybridization assays are described in WO92/02526 and U.S. Patent No. 5,124,246.

Alternatively, the Polymerase Chain Reaction (PCR) is another means for detecting small amounts of target nucleic acids (see, *e.g.*, Mullis et al., *Meth. Enzymol.* (1987) 155:335; U.S. Patent No. 4,683,195; and U.S. Patent No. 4,683,202). Two primer polynucleotides nucleotides that hybridize with the target nucleic acids are used to prime the reaction. The primers can be composed of sequence within or 3' and 5' to the polynucleotides of the Sequence Listing. Alternatively, if the primers are 3' and 5' to these

polynucleotides, they need not hybridize to them or the complements. After amplification of the target with a thermostable polymerase, the amplified target nucleic acids can be detected by methods known in the art, *e.g.*, Southern blot. mRNA or cDNA can also be detected by traditional blotting techniques (*e.g.*, Southern blot, Northern blot, *etc.*) described in Sambrook et al., "Molecular Cloning: A Laboratory Manual" (New York, Cold Spring Harbor Laboratory, 1989) (*e.g.*, without PCR amplification). In general, mRNA or cDNA generated from mRNA using a polymerase enzyme can be purified and separated using gel electrophoresis, and transferred to a solid support, such as nitrocellulose. The solid support is exposed to a labeled probe, washed to remove any unhybridized probe, and duplexes containing the labeled probe are detected.

Mapping. Polynucleotides of the present invention can be used to identify a chromosome on which the corresponding gene resides. Such mapping can be useful in identifying the function of the polynucleotide-related gene by its proximity to other genes with known function. Function can also be assigned to the polynucleotide-related gene when particular syndromes or diseases map to the same chromosome. For example, use of polynucleotide probes in identification and quantification of nucleic acid sequence aberrations is described in U.S. Patent No. 5,783,387. An exemplary mapping method is fluorescence *in situ* hybridization (FISH), which facilitates comparative genomic hybridization to allow total genome assessment of changes in relative copy number of DNA sequences (see, *e.g.*, Valdes et al., *Methods in Molecular Biology* (1997) 68:1). Polynucleotides can also be mapped to particular chromosomes using, for example, radiation hybrids or chromosome-specific hybrid panels. See Leach et al., *Advances in Genetics*, (1995) 33:63-99; Walter et al., *Nature Genetics* (1994) 7:22; Walter and Goodfellow, *Trends in Genetics* (1992) 9:352. Panels for radiation hybrid mapping are available from Research Genetics, Inc., Huntsville, Alabama, USA. The statistical program RHMAP can be used to construct a map based on the data from radiation hybridization with a measure of the relative likelihood of one order versus another. RHMAP is available via the world wide web at <http://www.sph.umich.edu/~group/statgen/software>. In addition, commercial programs are available for identifying regions of chromosomes commonly associated with disease, such as cancer.

Tissue Typing or Profiling. Expression of specific mRNA corresponding to the provided polynucleotides can vary in different cell types and can be tissue-specific. This variation of mRNA levels in different cell types can be exploited with nucleic acid probe assays to determine tissue types. For example, PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes substantially identical or complementary

to polynucleotides listed in the Sequence Listing can determine the presence or absence of the corresponding cDNA or mRNA.

Tissue typing can be used to identify the developmental organ or tissue source of a metastatic lesion by identifying the expression of a particular marker of that organ or tissue.

- 5 If a polynucleotide is expressed only in a specific tissue type, and a metastatic lesion is found to express that polynucleotide, then the developmental source of the lesion has been identified. Expression of a particular polynucleotide can be assayed by detection of either the corresponding mRNA or the protein product.

- 10 Use of Polymorphisms. A polynucleotide of the invention can be used in forensics, genetic analysis, mapping, and diagnostic applications where the corresponding region of a gene is polymorphic in the human population. Any means for detecting a polymorphism in a gene can be used, including, but not limited to electrophoresis of protein polymorphic variants, differential sensitivity to restriction enzyme cleavage, and hybridization to allele-specific probes.

#### 15 Antibody Production

- Expression products of a polynucleotide of the invention, as well as the corresponding mRNA, cDNA, or complete gene, can be prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. For polynucleotides to which a corresponding gene has not been assigned, this provides an additional method of  
20 identifying the corresponding gene. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native protein in a cell or tissue preparation or in a cell-free extract of an *in vitro* expression system.

- 25 Methods for production of monoclonal and polyclonal antibodies that specifically bind a selected antigen are well known in the art. The antibodies specifically bind to epitopes present in the polypeptides encoded by polynucleotides disclosed in the Sequence Listing. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. Epitopes that involve non-contiguous amino acids may require a longer  
30 polypeptide, *e.g.*, at least 15, 25, or 50 amino acids. Antibodies that specifically bind to human polypeptides encoded by the provided polynucleotides should provide a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in Western blots or other immunochemical assays. Preferably, antibodies that

specifically polypeptides of the invention do not bind to other proteins in immunochemical assays at detectable levels and can immunoprecipitate the specific polypeptide from solution.

The invention also contemplates naturally occurring antibodies specific for a polypeptide of the invention. For example, serum antibodies to a polypeptide of the invention in a human population can be purified by methods well known in the art, *e.g.*, by passing antiserum over a column to which the corresponding selected polypeptide or fusion protein is bound. The bound antibodies can then be eluted from the column, for example using a buffer with a high salt concentration.

In addition to the antibodies discussed above, the invention also contemplates genetically engineered antibodies, antibody derivatives (*e.g.*, single chain antibodies, antibody fragments (*e.g.*, Fab, *etc.*)), according to methods well known in the art.

Other embodiments of the present invention include humanized monoclonal antibodies capable of binding to the polypeptides of the invention. The phrase “humanized antibody” refers to an antibody derived from a non-human antibody - typically a mouse monoclonal antibody. Alternatively, a humanized antibody may be derived from a chimeric antibody that retains or substantially retains the antigen-binding properties of the parental, non-human, antibody but which exhibits diminished immunogenicity as compared to the parental antibody when administered to humans. The phrase “chimeric antibody,” as used herein, refers to an antibody containing sequence derived from two different antibodies (*see, e.g.*, U.S. Patent No. 4,816,567) which typically originate from different species. Most typically, chimeric antibodies comprise human and murine antibody fragments, generally human constant and mouse variable regions.

Because humanized antibodies are far less immunogenic in humans than the parental mouse monoclonal antibodies, they can be used for the treatment of humans with far less risk of anaphylaxis. Thus, these antibodies may be preferred in therapeutic applications that involve *in vivo* administration to a human such as, *e.g.*, use as radiation sensitizers for the treatment of neoplastic disease or use in methods to reduce the side effects of, *e.g.*, cancer therapy.

Humanized antibodies may be achieved by a variety of methods including, for example: (1) grafting the non-human complementarity determining regions (CDRs) onto a human framework and constant region (a process referred to in the art as “humanizing”), or, alternatively, (2) transplanting the entire non-human variable domains, but “cloaking” them with a human-like surface by replacement of surface residues (a process referred to in the art as “veneering”). In the present invention, humanized



antibodies will include both “humanized” and “veneered” antibodies. These methods are disclosed in, *e.g.*, Jones et al., *Nature* 321:522-525 (1986); Morrison et al., *Proc. Natl. Acad. Sci., U.S.A.*, 81:6851-6855 (1984); Morrison and Oi, *Adv. Immunol.*, 44:65-92 (1988); Verhoeyer et al., *Science* 239:1534-1536 (1988); Padlan, *Molec. Immunol.* 28:489-498 (1991); Padlan, *Molec. Immunol.* 31(3):169-217 (1994); and Kettleborough, C.A. et al., *Protein Eng.* 4(7):773-83 (1991) each of which is incorporated herein by reference.

The phrase “complementarity determining region” refers to amino acid sequences which together define the binding affinity and specificity of the natural Fv region of a native immunoglobulin binding site. *See, e.g.*, Chothia et al., *J. Mol. Biol.* 196:901-917 (1987); Kabat et al., U.S. Dept. of Health and Human Services NIH Publication No. 91-3242 (1991). The phrase “constant region” refers to the portion of the antibody molecule that confers effector functions. In the present invention, mouse constant regions are substituted by human constant regions. The constant regions of the subject humanized antibodies are derived from human immunoglobulins. The heavy chain constant region can be selected from any of the five isotypes: alpha, delta, epsilon, gamma or mu.

One method of humanizing antibodies comprises aligning the non-human heavy and light chain sequences to human heavy and light chain sequences, selecting and replacing the non-human framework with a human framework based on such alignment, molecular modeling to predict the conformation of the humanized sequence and comparing to the conformation of the parent antibody. This process is followed by repeated back mutation of residues in the CDR region which disturb the structure of the CDRs until the predicted conformation of the humanized sequence model closely approximates the conformation of the non-human CDRs of the parent non-human antibody. Such humanized antibodies may be further derivatized to facilitate uptake and clearance, *e.g.*, via Ashwell receptors. *See, e.g.*, U.S. Patent Nos. 5,530,101 and 5,585,089 which patents are incorporated herein by reference.

Humanized antibodies can also be produced using transgenic animals that are engineered to contain human immunoglobulin loci. For example, WO 98/24893 discloses transgenic animals having a human Ig locus wherein the animals do not produce functional endogenous immunoglobulins due to the inactivation of endogenous heavy and light chain loci. WO 91/10741 also discloses transgenic non-primate mammalian hosts capable of mounting an immune response to an immunogen, wherein the antibodies have primate constant and/or variable regions, and wherein the endogenous immunoglobulin-encoding loci are substituted or inactivated. WO 96/30498 discloses the use of the Cre/Lox

system to modify the immunoglobulin locus in a mammal, such as to replace all or a portion of the constant or variable region to form a modified antibody molecule. WO 94/02602 discloses non-human mammalian hosts having inactivated endogenous Ig loci and functional human Ig loci. U.S. Patent No. 5,939,598 discloses methods of making  
 5 transgenic mice in which the mice lack endogenous heavy chains, and express an exogenous immunoglobulin locus comprising one or more xenogeneic constant regions.

Using a transgenic animal described above, an immune response can be produced to a selected antigenic molecule, and antibody-producing cells can be removed from the animal and used to produce hybridomas that secrete human monoclonal  
 10 antibodies. Immunization protocols, adjuvants, and the like are known in the art, and are used in immunization of, for example, a transgenic mouse as described in WO 96/33735. This publication discloses monoclonal antibodies against a variety of antigenic molecules including IL-6, IL-8, TNF $\alpha$ , human CD4, L-selectin, gp39, and tetanus toxin. The monoclonal antibodies can be tested for the ability to inhibit or neutralize the biological  
 15 activity or physiological effect of the corresponding protein. WO 96/33735 discloses that monoclonal antibodies against IL-8, derived from immune cells of transgenic mice immunized with IL-8, blocked IL-8-induced functions of neutrophils. Human monoclonal antibodies with specificity for the antigen used to immunize transgenic animals are also disclosed in WO 96/34096.

#### Polynucleotides or Arrays for Diagnostics

Polynucleotide arrays are created by spotting polynucleotide probes onto a substrate (e.g., glass, nitrocellulose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific  
 25 interactions, such as hydrophobic interactions. Samples of polynucleotides can be detectably labeled (e.g., using radioactive or fluorescent labels) and then hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Techniques for constructing arrays and methods of using these arrays are  
 30 described in EP 799 897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; U.S. Patent No. 5,593,839; U.S. Patent No. 5,578,832; EP 728 520; U.S. Patent No. 5,599,695; EP 721 016; U.S. Patent No. 5,556,752; WO 95/22058; and U.S. Patent No. 5,631,734. Arrays can be used to, for example, examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential

expression of a polynucleotide between a test cell and control cell (*e.g.*, cancer cells and normal cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific gene product. Exemplary uses of arrays are further described in, for example, Pappalarado et al., *Sem. Radiation Oncol.* (1998) 8:217; and Ramsay, *Nature Biotechnol.* (1998) 16:40.

#### Differential Expression in Diagnosis

The polynucleotides of the invention can also be used to detect differences in expression levels between two cells, *e.g.*, as a method to identify abnormal or diseased tissue in a human. For polynucleotides corresponding to profiles of protein families, the choice of tissue can be selected according to the putative biological function. In general, the expression of a gene corresponding to a specific polynucleotide is compared between a first tissue that is suspected of being diseased and a second, normal tissue of the human. The tissue suspected of being abnormal or diseased can be derived from a different tissue type of the human, but preferably it is derived from the same tissue type; for example an intestinal polyp or other abnormal growth should be compared with normal intestinal tissue. The normal tissue can be the same tissue as that of the test sample, or any normal tissue of the patient, especially those that express the polynucleotide-related gene of interest (*e.g.*, brain, thymus, testis, heart, prostate, placenta, spleen, small intestine, skeletal muscle, pancreas, and the mucosal lining of the colon). A difference between the polynucleotide-related gene, mRNA, or protein in the two tissues which are compared, for example in molecular weight, amino acid or nucleotide sequence, or relative abundance, indicates a change in the gene, or a gene which regulates it, in the tissue of the human that was suspected of being diseased. Examples of detection of differential expression and its use in diagnosis of cancer are described in U.S. Patent Nos. 5,688,641 and 5,677,125.

A genetic predisposition to disease in a human can also be detected by comparing expression levels of an mRNA or protein corresponding to a polynucleotide of the invention in a fetal tissue with levels associated in normal fetal tissue. Fetal tissues that are used for this purpose include, but are not limited to, amniotic fluid, chorionic villi, blood, and the blastomere of an *in vitro*-fertilized embryo. The comparable normal polynucleotide-related gene is obtained from any tissue. The mRNA or protein is obtained from a normal tissue of a human in which the polynucleotide-related gene is expressed. Differences such as alterations in the nucleotide sequence or size of the same product of the fetal polynucleotide-related gene or mRNA, or alterations in the molecular weight, amino

acid sequence, or relative abundance of fetal protein, can indicate a germline mutation in the polynucleotide-related gene of the fetus, which indicates a genetic predisposition to disease. In general, diagnostic, prognostic, and other methods of the invention based on differential expression involve detection of a level or amount of a gene product, particularly a differentially expressed gene product, in a test sample obtained from a patient suspected of having or being susceptible to a disease (*e.g.*, breast cancer, lung cancer, colon cancer and/or metastatic forms thereof), and comparing the detected levels to those levels found in normal cells (*e.g.*, cells substantially unaffected by cancer) and/or other control cells (*e.g.*, to differentiate a cancerous cell from a cell affected by dysplasia). Furthermore, the severity of the disease can be assessed by comparing the detected levels of a differentially expressed gene product with those levels detected in samples representing the levels of differentially gene product associated with varying degrees of severity of disease. It should be noted that use of the term “diagnostic” herein is not necessarily meant to exclude “prognostic” or “prognosis,” but rather is used as a matter of convenience.

The term “differentially expressed gene” is generally intended to encompass a polynucleotide that can, for example, include an open reading frame encoding a gene product (*e.g.*, a polypeptide), and/or introns of such genes and adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene can be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome. In general, a difference in expression level associated with a decrease in expression level of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% or more is indicative of a differentially expressed gene of interest, *i.e.*, a gene that is underexpressed or down-regulated in the test sample relative to a control sample. Furthermore, a difference in expression level associated with an increase in expression of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% and can be at least about 1 ½-fold, usually at least about 2-fold to about 10-fold, and can be about 100-fold to about 1,000-fold increase relative to a control sample is indicative of a differentially expressed gene of interest, *i.e.*, an overexpressed or up-regulated gene.

“Differentially expressed polynucleotide” as used herein means a nucleic acid molecule (RNA or DNA) comprising a sequence that represents a differentially expressed gene, *e.g.*, the differentially expressed polynucleotide comprises a sequence (*e.g.*, an open reading frame encoding a gene product) that uniquely identifies a differentially expressed gene so that detection of the differentially expressed polynucleotide in a sample is

correlated with the presence of a differentially expressed gene in a sample. "Differentially expressed polynucleotides" is also meant to encompass fragments of the disclosed polynucleotides, *e.g.*, fragments retaining biological activity, as well as nucleic acids homologous, substantially similar, or substantially identical (*e.g.*, having about 90% sequence identity) to the disclosed polynucleotides.

"Diagnosis" as used herein generally includes determination of a subject's susceptibility to a disease or disorder, determination as to whether a subject is presently affected by a disease or disorder, as well as to the prognosis of a subject affected by a disease or disorder (*e.g.*, identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). The present invention particularly encompasses diagnosis of subjects in the context of breast cancer (*e.g.*, carcinoma *in situ* (*e.g.*, ductal carcinoma *in situ*), estrogen receptor (ER)-positive breast cancer, ER-negative breast cancer, or other forms and/or stages of breast cancer), lung cancer (*e.g.*, small cell carcinoma, non-small cell carcinoma, mesothelioma, and other forms and/or stages of lung cancer), and colon cancer (*e.g.*, adenomatous polyp, colorectal carcinoma, and other forms and/or stages of colon cancer).

"Sample" or "biological sample" as used throughout here are generally meant to refer to samples of biological fluids or tissues, particularly samples obtained from tissues, especially from cells of the type associated with the disease for which the diagnostic application is designed (*e.g.*, ductal adenocarcinoma), and the like. "Samples" is also meant to encompass derivatives and fractions of such samples (*e.g.*, cell lysates). Where the sample is solid tissue, the cells of the tissue can be dissociated or tissue sections can be analyzed.

Methods of the subject invention useful in diagnosis or prognosis typically involve comparison of the abundance of a selected differentially expressed gene product in a sample of interest with that of a control to determine any relative differences in the expression of the gene product, where the difference can be measured qualitatively and/or quantitatively. Quantitation can be accomplished, for example, by comparing the level of expression product detected in the sample with the amounts of product present in a standard curve. A comparison can be made visually; by using a technique such as densitometry, with or without computerized assistance; by preparing a representative library of cDNA clones of mRNA isolated from a test sample, sequencing the clones in the library to determine that number of cDNA clones corresponding to the same gene product, and analyzing the number of clones corresponding to that same gene product relative to the number of clones of the same gene product in a control sample; or by using an array to

detect relative levels of hybridization to a selected sequence or set of sequences, and comparing the hybridization pattern to that of a control. The differences in expression are then correlated with the presence or absence of an abnormal expression pattern. A variety of different methods for determining the nucleic acid abundance in a sample are known to those of skill in the art (see, *e.g.*, WO 97/27317). In general, diagnostic assays of the invention involve detection of a gene product of a the polynucleotide sequence (*e.g.*, mRNA or polypeptide) that corresponds to a sequence of SEQ ID NOs:1-3351. The patient from whom the sample is obtained can be apparently healthy, susceptible to disease (*e.g.*, as determined by family history or exposure to certain environmental factors), or can already be identified as having a condition in which altered expression of a gene product of the invention is implicated.

Diagnosis can be determined based on detected gene product expression levels of a gene product encoded by at least one, preferably at least two or more, at least 3 or more, or at least 4 or more of the polynucleotides having a sequence set forth in SEQ ID NOs:1-3351, and can involve detection of expression of genes corresponding to all of SEQ ID NOs:1-3351 and/or additional sequences that can serve as additional diagnostic markers and/or reference sequences. Where the diagnostic method is designed to detect the presence or susceptibility of a patient to cancer, the assay preferably involves detection of a gene product encoded by a gene corresponding to a polynucleotide that is differentially expressed in cancer. Examples of such differentially expressed polynucleotides are described in the Examples below. Given the provided polynucleotides and information regarding their relative expression levels provided herein, assays using such polynucleotides and detection of their expression levels in diagnosis and prognosis will be readily apparent to the ordinarily skilled artisan.

Any of a variety of detectable labels can be used in connection with the various embodiments of the diagnostic methods of the invention. Suitable detectable labels include fluorochromes, (*e.g.*, fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)), radioactive labels, (*e.g.*,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ , *etc.*), and the like. The detectable label can involve a two stage systems (*e.g.*, biotin-avidin, hapten-anti-hapten antibody, *etc.*)

Reagents specific for the polynucleotides and polypeptides of the invention, such as antibodies and nucleotide probes, can be supplied in a kit for detecting the presence of an

expression product in a biological sample. The kit can also contain buffers or labeling components, as well as instructions for using the reagents to detect and quantify expression products in the biological sample. Exemplary embodiments of the diagnostic methods of the invention are described below in more detail.

5        Polypeptide detection in diagnosis. In one embodiment, the test sample is assayed for the level of a differentially expressed polypeptide. Diagnosis can be accomplished using any of a number of methods to determine the absence or presence or altered amounts of the differentially expressed polypeptide in the test sample. For example, detection can utilize staining of cells or histological sections with labeled antibodies, performed in  
10 accordance with conventional methods. Cells can be permeabilized to stain cytoplasmic molecules. In general, antibodies that specifically bind a differentially expressed polypeptide of the invention are added to a sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody can be detectably labeled for direct detection (*e.g.*, using radioisotopes, enzymes,  
15 fluorescers, chemiluminescers, and the like), or can be used in conjunction with a second stage antibody or reagent to detect binding (*e.g.*, biotin with horseradish peroxidase-conjugated avidin, a secondary antibody conjugated to a fluorescent compound, *e.g.*, fluorescein, rhodamine, Texas red, *etc.*). The absence or presence of antibody binding can be determined by various methods, including flow cytometry of dissociated cells,  
20 microscopy, radiography, scintillation counting, *etc.* Any suitable alternative methods can of qualitative or quantitative detection of levels or amounts of differentially expressed polypeptide can be used, for example ELISA, western blot, immunoprecipitation, radioimmunoassay, *etc.*

mRNA detection. The diagnostic methods of the invention can also or alternatively  
25 involve detection of mRNA encoded by a gene corresponding to a differentially expressed polynucleotides of the invention. Any suitable qualitative or quantitative methods known in the art for detecting specific mRNAs can be used. mRNA can be detected by, for example, *in situ* hybridization in tissue sections, by reverse transcriptase-PCR, or in Northern blots containing poly A+ mRNA. One of skill in the art can readily use these  
30 methods to determine differences in the size or amount of mRNA transcripts between two samples. mRNA expression levels in a sample can also be determined by generation of a library of expressed sequence tags (ESTs) from the sample, where the EST library is representative of sequences present in the sample (Adams, et al., (1991) Science 252:1651). Enumeration of the relative representation of ESTs within the library can be used to  
35 approximate the relative representation of the gene transcript within the starting sample.

The results of EST analysis of a test sample can then be compared to EST analysis of a reference sample to determine the relative expression levels of a selected polynucleotide, particularly a polynucleotide corresponding to one or more of the differentially expressed genes described herein. Alternatively, gene expression in a test sample can be performed using serial analysis of gene expression (SAGE) methodology (*e.g.*, Velculescu et al., *Science* (1995) 270:484) or differential display (DD) methodology (see, *e.g.*, U.S. Patent NOs. 5,776,683 and 5,807,680).

Alternatively, gene expression can be analyzed using hybridization analysis. Oligonucleotides or cDNA can be used to selectively identify or capture DNA or RNA of specific sequence composition, and the amount of RNA or cDNA hybridized to a known capture sequence determined qualitatively or quantitatively, to provide information about the relative representation of a particular message within the pool of cellular messages in a sample. Hybridization analysis can be designed to allow for concurrent screening of the relative expression of hundreds to thousands of genes by using, for example, array-based technologies having high density formats, including filters, microscope slides, or microchips, or solution-based technologies that use spectroscopic analysis (*e.g.*, mass spectrometry). One exemplary use of arrays in the diagnostic methods of the invention is described below in more detail.

Use of a single gene in diagnostic applications. The diagnostic methods of the invention can focus on the expression of a single differentially expressed gene. For example, the diagnostic method can involve detecting a differentially expressed gene, or a polymorphism of such a gene (*e.g.*, a polymorphism in an coding region or control region), that is associated with disease. Disease-associated polymorphisms can include deletion or truncation of the gene, mutations that alter expression level and/or affect activity of the encoded protein, *etc.*

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, *e.g.*, a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express a differentially expressed gene can be used as a source of mRNA, which can be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid can be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis, and a detectable label can be included in the amplification reaction (*e.g.*, using a detectably labeled primer or detectably labeled oligonucleotides) to facilitate detection. Alternatively, various methods are also known in the art that utilize



oligonucleotide ligation as a means of detecting polymorphisms, see *e.g.*, Riley et al., *Nucl. Acids Res.* (1990) 18:2887; and Delahunty et al., *Am. J. Hum. Genet.* (1996) 58:1239.

The amplified or cloned sample nucleic acid can be analyzed by one of a number of methods known in the art. The nucleic acid can be sequenced by dideoxy or other methods, and the sequence of bases compared to a selected sequence, *e.g.*, to a wild-type sequence. Hybridization with the polymorphic or variant sequence can also be used to determine its presence in a sample (*e.g.*, by Southern blot, dot blot, *etc.*). The hybridization pattern of a polymorphic or variant sequence and a control sequence to an array of oligonucleotide probes immobilized on a solid support, as described in U.S. Patent No. 5,445,934, or in WO 95/35505, can also be used as a means of identifying polymorphic or variant sequences associated with disease. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

Screening for mutations in a gene can be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that can affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in proteins can be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded protein can be determined by comparison with the wild-type protein.

Pattern matching in diagnosis using arrays. In another embodiment, the diagnostic and/or prognostic methods of the invention involve detection of expression of a selected set of genes in a test sample to produce a test expression pattern (TEP). The TEP is compared to a reference expression pattern (REP), which is generated by detection of expression of the selected set of genes in a reference sample (*e.g.*, a positive or negative control sample). The selected set of genes includes at least one of the genes of the invention, which genes correspond to the polynucleotide sequences of SEQ ID NOs:1-3351. Of particular interest is a selected set of genes that includes genes differentially expressed in the disease for which the test sample is to be screened.

“Reference sequences” or “reference polynucleotides” as used herein in the context of differential gene expression analysis and diagnosis/prognosis refers to a selected set of polynucleotides, which selected set includes at least one or more of the differentially expressed polynucleotides described herein. A plurality of reference sequences, preferably comprising positive and negative control sequences, can be included as reference sequences. Additional suitable reference sequences are found in Genbank, Unigene, and other nucleotide sequence databases (including, *e.g.*, expressed sequence tag (EST), partial, and full-length sequences).

“Reference array” means an array having reference sequences for use in hybridization with a sample, where the reference sequences include all, at least one of, or any subset of the differentially expressed polynucleotides described herein. Usually such an array will include at least 3 different reference sequences, and can include any one or all of the provided differentially expressed sequences. Arrays of interest can further comprise sequences, including polymorphisms, of other genetic sequences, particularly other sequences of interest for screening for a disease or disorder (*e.g.*, cancer, dysplasia, or other related or unrelated diseases, disorders, or conditions). The oligonucleotide sequence on the array will usually be at least about 12 nt in length, and can be of about the length of the provided sequences, or can extend into the flanking regions to generate fragments of 100 nt to 200 nt in length or more. Reference arrays can be produced according to any suitable methods known in the art. For example, methods of producing large arrays of oligonucleotides are described in U.S. Patent NOs. 5,134,854 and 5,445,934 using light-directed synthesis techniques. Using a computer controlled system, a heterogeneous array of monomers is converted, through simultaneous coupling at a number of reaction sites, into a heterogeneous array of polymers. Alternatively, microarrays are generated by deposition of pre-synthesized oligonucleotides onto a solid substrate, for example as described in PCT published application no. WO 95/35505.

A “reference expression pattern” or “REP” as used herein refers to the relative levels of expression of a selected set of genes, particularly of differentially expressed genes, that is associated with a selected cell type, *e.g.*, a normal cell, a cancerous cell, a cell exposed to an environmental stimulus, and the like. A “test expression pattern” or “TEP” refers to relative levels of expression of a selected set of genes, particularly of differentially expressed genes, in a test sample (*e.g.*, a cell of unknown or suspected disease state, from which mRNA is isolated).

REPs can be generated in a variety of ways according to methods well known in the art. For example, REPs can be generated by hybridizing a control sample to an array

having a selected set of polynucleotides (particularly a selected set of differentially expressed polynucleotides), acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the REP with a TEP. Alternatively, all expressed sequences in a control sample can be isolated and sequenced, *e.g.*, by isolating mRNA from a control sample, converting the mRNA into cDNA, and sequencing the cDNA. The resulting sequence information roughly or precisely reflects the identity and relative number of expressed sequences in the sample. The sequence information can then be stored in a format (*e.g.*, a computer-readable format) that allows for ready comparison of the REP with a TEP. The REP can be normalized prior to or after data storage, and/or can be processed to selectively remove sequences of expressed genes that are of less interest or that might complicate analysis (*e.g.*, some or all of the sequences associated with housekeeping genes can be eliminated from REP data).

TEPs can be generated in a manner similar to REPs, *e.g.*, by hybridizing a test sample to an array having a selected set of polynucleotides, particularly a selected set of differentially expressed polynucleotides, acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the TEP with a REP. The REP and TEP to be used in a comparison can be generated simultaneously, or the TEP can be compared to previously generated and stored REPs.

In one embodiment of the invention, comparison of a TEP with a REP involves hybridizing a test sample with a reference array, where the reference array has one or more reference sequences for use in hybridization with a sample. The reference sequences include all, at least one of, or any subset of the differentially expressed polynucleotides described herein. Hybridization data for the test sample is acquired, the data normalized, and the produced TEP compared with a REP generated using an array having the same or similar selected set of differentially expressed polynucleotides. Probes that correspond to sequences differentially expressed between the two samples will show decreased or increased hybridization efficiency for one of the samples relative to the other.

Methods for collection of data from hybridization of samples with a reference arrays are well known in the art. For example, the polynucleotides of the reference and test samples can be generated using a detectable fluorescent label, and hybridization of the polynucleotides in the samples detected by scanning the microarrays for the presence of the detectable label using, for example, a microscope and light source for directing light at a substrate. A photon counter detects fluorescence from the substrate, while an x-y translation stage varies the location of the substrate. A confocal detection device that can be used in the subject methods is described in U.S. Patent No. 5,631,734. A scanning laser

microscope is described in Shalon et al., *Genome Res.* (1996) 6:639. A scan, using the appropriate excitation line, is performed for each fluorophore used. The digital images generated from the scan are then combined for subsequent analysis. For any particular array element, the ratio of the fluorescent signal from one sample (*e.g.*, a test sample) is compared to the fluorescent signal from another sample (*e.g.*, a reference sample), and the relative signal intensity determined.

Methods for analyzing the data collected from hybridization to arrays are well known in the art. For example, where detection of hybridization involves a fluorescent label, data analysis can include the steps of determining fluorescent intensity as a function of substrate position from the data collected, removing outliers, *i.e.*, data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the targets from the remaining data. The resulting data can be displayed as an image with the intensity in each region varying according to the binding affinity between targets and probes.

In general, the test sample is classified as having a gene expression profile corresponding to that associated with a disease or non-disease state by comparing the TEP generated from the test sample to one or more REPs generated from reference samples (*e.g.*, from samples associated with cancer or specific stages of cancer, dysplasia, samples affected by a disease other than cancer, normal samples, *etc.*). The criteria for a match or a substantial match between a TEP and a REP include expression of the same or substantially the same set of reference genes, as well as expression of these reference genes at substantially the same levels (*e.g.*, no significant difference between the samples for a signal associated with a selected reference sequence after normalization of the samples, or at least no greater than about 25% to about 40% difference in signal strength for a given reference sequence. In general, a pattern match between a TEP and a REP includes a match in expression, preferably a match in qualitative or quantitative expression level, of at least one of, all or any subset of the differentially expressed genes of the invention.

Pattern matching can be performed manually, or can be performed using a computer program. Methods for preparation of substrate matrices (*e.g.*, arrays), design of oligonucleotides for use with such matrices, labeling of probes, hybridization conditions, scanning of hybridized matrices, and analysis of patterns generated, including comparison analysis, are described in, for example, U.S. Patent No. 5,800,992.

### Diagnosis, Prognosis and Management of Cancer

The polynucleotides of the invention and their gene products are of particular interest as genetic or biochemical markers (*e.g.*, in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or to monitor the efficacy of various therapies and preventive interventions. For example, the level of expression of certain polynucleotides can be indicative of a poorer prognosis, and therefore warrant more aggressive chemo- or radio-therapy for a patient or vice versa. The correlation of novel surrogate tumor specific features with response to treatment and outcome in patients can define prognostic indicators that allow the design of tailored therapy based on the molecular profile of the tumor. These therapies include antibody targeting and gene therapy. Determining expression of certain polynucleotides and comparison of a patient's profile with known expression in normal tissue and variants of the disease allows a determination of the best possible treatment for a patient, both in terms of specificity of treatment and in terms of comfort level of the patient. Surrogate tumor markers, such as polynucleotide expression, can also be used to better classify, and thus diagnose and treat, different forms and disease states of cancer. Two classifications widely used in oncology that can benefit from identification of the expression levels of the polynucleotides of the invention are staging of the cancerous disorder, and grading the nature of the cancerous tissue.

The polynucleotides of the invention can be useful to monitor patients having or susceptible to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. Furthermore, a polynucleotide of the invention identified as important for one type of cancer can also have implications for development or risk of development of other types of cancer, *e.g.*, where a polynucleotide is differentially expressed across various cancer types. Thus, for example, expression of a polynucleotide that has clinical implications for metastatic colon cancer can also have clinical implications for stomach cancer or endometrial cancer.

Staging. Staging is a process used by physicians to describe how advanced the cancerous state is in a patient. Generally, if a cancer is only detectable in the area of the primary lesion without having spread to any lymph nodes it is called Stage I. If it has spread only to the closest lymph nodes, it is called Stage II. In Stage III, the cancer has generally spread to the lymph nodes in near proximity to the site of the primary lesion. Cancers that have spread to a distant part of the body, such as the liver, bone, brain or other site, are Stage IV, the most advanced stage.

The polynucleotides of the invention can facilitate fine-tuning of the staging process by identifying markers for the aggressivity of a cancer, *e.g.*, the metastatic potential, as well as the presence in different areas of the body. Thus, a Stage II cancer with a polynucleotide signifying a high metastatic potential cancer can be used to change a borderline Stage II tumor to a Stage III tumor, justifying more aggressive therapy. Conversely, the presence of a polynucleotide signifying a lower metastatic potential allows more conservative staging of a tumor.

Grading of cancers. Grade is a term used to describe how closely a tumor resembles normal tissue of its same type. The microscopic appearance of a tumor is used to identify tumor grade based on parameters such as cell morphology, cellular organization, and other markers of differentiation. As a general rule, the grade of a tumor corresponds to its rate of growth or aggressiveness, with undifferentiated or high-grade tumors being more aggressive than well differentiated or low-grade tumors. The following guidelines are generally used for grading tumors: 1) GX Grade cannot be assessed; 2) G1 Well differentiated; G2 Moderately well differentiated; 3) G3 Poorly differentiated; 4) G4 Undifferentiated. The polynucleotides of the invention can be especially valuable in determining the grade of the tumor, as they not only can aid in determining the differentiation status of the cells of a tumor, they can also identify factors other than differentiation that are valuable in determining the aggressivity of a tumor, such as metastatic potential.

Detection of lung cancer. The polynucleotides of the invention can be used to detect lung cancer in a subject. Although there are more than a dozen different kinds of lung cancer, the two main types of lung cancer are small cell and nonsmall cell, which encompass about 90% of all lung cancer cases. Small cell carcinoma (also called oat cell carcinoma) usually starts in one of the larger bronchial tubes, grows fairly rapidly, and is likely to be large by the time of diagnosis. Nonsmall cell lung cancer (NSCLC) is made up of three general subtypes of lung cancer. Epidermoid carcinoma (also called squamous cell carcinoma) usually starts in one of the larger bronchial tubes and grows relatively slowly. The size of these tumors can range from very small to quite large. Adenocarcinoma starts growing near the outside surface of the lung and can vary in both size and growth rate. Some slowly growing adenocarcinomas are described as alveolar cell cancer. Large cell carcinoma starts near the surface of the lung, grows rapidly, and the growth is usually fairly large when diagnosed. Other less common forms of lung cancer are carcinoid, cylindroma, mucoepidermoid, and malignant mesothelioma.

The polynucleotides of the invention, *e.g.*, polynucleotides differentially expressed in normal cells versus cancerous lung cells (*e.g.*, tumor cells of high or low metastatic potential) or between types of cancerous lung cells (*e.g.*, high metastatic versus low metastatic), can be used to distinguish types of lung cancer as well as identifying traits specific to a certain patient's cancer and selecting an appropriate therapy. For example, if the patient's biopsy expresses a polynucleotide that is associated with a low metastatic potential, it may justify leaving a larger portion of the patient's lung in surgery to remove the lesion. Alternatively, a smaller lesion with expression of a polynucleotide that is associated with high metastatic potential may justify a more radical removal of lung tissue and/or the surrounding lymph nodes, even if no metastasis can be identified through pathological examination.

Detection of breast cancer. The majority of breast cancers are adenocarcinomas subtypes, which can be summarized as follows: 1) ductal carcinoma *in situ* (DCIS), including comedocarcinoma; 2) infiltrating (or invasive) ductal carcinoma (IDC); 3) lobular carcinoma *in situ* (LCIS); 4) infiltrating (or invasive) lobular carcinoma (ILC); 5) inflammatory breast cancer; 6) medullary carcinoma; 7) mucinous carcinoma; 8) Paget's disease of the nipple; 9) Phyllodes tumor; and 10) tubular carcinoma.

The expression of polynucleotides of the invention can be used in the diagnosis and management of breast cancer, as well as to distinguish between types of breast cancer. Detection of breast cancer can be determined using expression levels of any of the appropriate polynucleotides of the invention, either alone or in combination. Determination of the aggressive nature and/or the metastatic potential of a breast cancer can also be determined by comparing levels of one or more polynucleotides of the invention and comparing levels of another sequence known to vary in cancerous tissue, *e.g.*, ER expression. In addition, development of breast cancer can be detected by examining the ratio of expression of a differentially expressed polynucleotide to the levels of steroid hormones (*e.g.*, testosterone or estrogen) or to other hormones (*e.g.*, growth hormone, insulin). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous breast tissue, to discriminate between breast cancers with different cells of origin, to discriminate between breast cancers with different potential metastatic rates, *etc.*

Detection of colon cancer. The polynucleotides of the invention exhibiting the appropriate expression pattern can be used to detect colon cancer in a subject. Colorectal cancer is one of the most common neoplasms in humans and perhaps the most frequent form of hereditary neoplasia. Prevention and early detection are key factors in controlling

and curing colorectal cancer. Colorectal cancer begins as polyps, which are small, benign growths of cells that form on the inner lining of the colon. Over a period of several years, some of these polyps accumulate additional mutations and become cancerous. Multiple familial colorectal cancer disorders have been identified, which are summarized as follows:

5 1) Familial adenomatous polyposis (FAP); 2) Gardner's syndrome; 3) Hereditary nonpolyposis colon cancer (HNPCC); and 4) Familial colorectal cancer in Ashkenazi Jews. The expression of appropriate polynucleotides of the invention can be used in the diagnosis, prognosis and management of colorectal cancer. Detection of colon cancer can be determined using expression levels of any of these sequences alone or in combination

10 with the levels of expression. Determination of the aggressive nature and/or the metastatic potential of a colon cancer can be determined by comparing levels of one or more polynucleotides of the invention and comparing total levels of another sequence known to vary in cancerous tissue, *e.g.*, expression of p53, DCC, ras, or FAP (see, *e.g.*, Fearon ER, et al., *Cell* (1990) 61(5):759; Hamilton SR et al., *Cancer* (1993) 72:957; Bodmer W, et al.,

15 *Nat Genet.* (1994) 4(3):217; Fearon ER, *Ann N Y Acad Sci.* (1995) 768:101). For example, development of colon cancer can be detected by examining the ratio of any of the polynucleotides of the invention to the levels of oncogenes (*e.g.*, ras) or tumor suppressor genes (*e.g.*, FAP or p53). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous colon tissue, to discriminate between colon

20 cancers with different cells of origin, to discriminate between colon cancers with different potential metastatic rates, *etc.*

#### Use of Polynucleotides to Screen for Peptide Analogs and Antagonists

Polypeptides encoded by the instant polynucleotides and corresponding full length genes can be used to screen peptide libraries to identify binding partners, such as receptors,

25 from among the encoded polypeptides. Peptide libraries can be synthesized according to methods known in the art (see, *e.g.*, U.S. Patent No. 5,010,175, and WO 91/17823). Agonists or antagonists of the polypeptides if the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, *etc.* The assay conditions ideally should

30 resemble the conditions under which the native activity is exhibited *in vivo*, that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to



the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

Such screening and experimentation can lead to identification of a novel polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide of the invention, and at least one peptide agonist or antagonist of the novel binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the novel receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

#### Pharmaceutical Compositions and Therapeutic Uses

Pharmaceutical compositions of the invention can comprise polypeptides, antibodies, or polynucleotides (including antisense nucleotides and ribozymes) of the claimed invention in a therapeutically effective amount. The term “therapeutically effective amount” as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term “pharmaceutically acceptable carrier” refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which

can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

5 Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid

10 vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, *e.g.*, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough

15 discussion of pharmaceutically acceptable excipients is available in *Remington's Pharmaceutical Sciences* (Mack Pub. Co., New Jersey, 1991).

Delivery Methods. Once formulated, the compositions of the invention can be (1) administered directly to the subject (*e.g.*, as polynucleotide or polypeptides); or (2) delivered *ex vivo*, to cells derived from the subject (*e.g.*, as in *ex vivo* gene therapy).

20 Direct delivery of the compositions will generally be accomplished by parenteral injection, *e.g.*, subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumoral or to the interstitial space of a tissue. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.

25 Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *e.g.*, International Publication No. WO 93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be

30 accomplished by, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Once a gene corresponding to a polynucleotide of the invention has been found to

35 correlate with a proliferative disorder, such as neoplasia, dysplasia, and hyperplasia, the

disorder can be amenable to treatment by administration of a therapeutic agent based on the provided polynucleotide, corresponding polypeptide or other corresponding molecule (e.g., antisense, ribozyme, etc.).

The dose and the means of administration of the inventive pharmaceutical compositions are determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. For example, administration of polynucleotide therapeutic compositions agents of the invention includes local or systemic administration, including injection, oral administration, particle gun or catheterized administration, and topical administration. Preferably, the therapeutic polynucleotide composition contains an expression construct comprising a promoter operably linked to a polynucleotide of at least 12, 22, 25, 30, or 35 contiguous nt of the polynucleotide disclosed herein. Various methods can be used to administer the therapeutic composition directly to a specific site in the body. For example, a small metastatic lesion is located and the therapeutic composition injected several times in several different locations within the body of tumor. Alternatively, arteries which serve a tumor are identified, and the therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor. A tumor that has a necrotic center is aspirated and the composition injected directly into the now empty center of the tumor. The antisense composition is directly administered to the surface of the tumor, for example, by topical application of the composition. X-ray imaging is used to assist in certain of the above delivery methods.

Receptor-mediated targeted delivery of therapeutic compositions containing an antisense polynucleotide, subgenomic polynucleotides, or antibodies to specific tissues can also be used. Receptor-mediated DNA delivery techniques are described in, for example, Findeis et al., *Trends Biotechnol.* (1993) 11:202; Chiou et al., *Gene Therapeutics: Methods And Applications Of Direct Gene Transfer* (J.A. Wolff, ed.) (1994); Wu et al., *J. Biol. Chem.* (1988) 263:621; Wu et al., *J. Biol. Chem.* (1994) 269:542; Zenke et al., *Proc. Natl. Acad. Sci. (USA)* (1990) 87:3655; Wu et al., *J. Biol. Chem.* (1991) 266:338. Therapeutic compositions containing a polynucleotide are administered in a range of about 100 ng to about 200 mg of DNA for local administration in a gene therapy protocol. Concentration ranges of about 500 ng to about 50 mg, about 1 mg to about 2 mg, about 5 mg to about 500 mg, and about 20 mg to about 100 mg of DNA can also be used during a gene therapy protocol. Factors such as method of action (e.g., for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations which will affect the dosage required for ultimate efficacy of the antisense subgenomic

polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of antisense subgenomic polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of, for example, a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect. For polynucleotide-related genes encoding polypeptides or proteins with anti-inflammatory activity, suitable use, doses, and administration are described in U.S. Patent No. 5,654,173.

The therapeutic polynucleotides and polypeptides of the present invention can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally, Jolly, *Cancer Gene Therapy* (1994) 1:51; Kimura, *Human Gene Therapy* (1994) 5:845; Connelly, *Human Gene Therapy* (1995) 1:185; and Kaplitt, *Nature Genetics* (1994) 6:148). Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

Viral-based vectors for delivery of a desired polynucleotide and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (see, e.g., WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; U.S. Patent No. 5, 219,740; WO 93/11230; WO 93/10218; U.S. Patent No. 4,777,127; GB Patent No. 2,200,651; EP 0 345 242; and WO 91/02805), alphavirus-based vectors (e.g., Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), and adeno-associated virus (AAV) vectors (see, e.g., WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655). Administration of DNA linked to killed adenovirus as described in Curiel, *Hum. Gene Ther.* (1992) 3:147 can also be employed.

Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone (see, e.g., Curiel, *Hum. Gene Ther.* (1992) 3:147); ligand-linked DNA (see, e.g., Wu, *J. Biol. Chem.* 264:16985 (1989)); eukaryotic cell delivery vehicles (see, e.g., U.S. Patent No. 5,814,482; WO 95/07994; WO 96/17072; WO 95/30763; and WO 97/42338) and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and U.S. Patent No. 5,580,859. Liposomes that can act as gene delivery vehicles are described in

U.S. Patent No. 5,422,120; WO 95/13796; WO 94/23697; WO 91/14445; and EP 0524968. Additional approaches are described in Philip, *Mol. Cell Biol.* 14:2411 (1994), and in Woffendin, *Proc. Natl. Acad. Sci.* (1994) 91:1581.

Further non-viral delivery suitable for use includes mechanical delivery systems  
5 such as the approach described in Woffendin et al., *Proc. Natl. Acad. Sci. USA*  
91(24):11581 (1994). Moreover, the coding sequence and the product of expression of  
such can be delivered through deposition of photopolymerized hydrogel materials or use of  
ionizing radiation (see, e.g., U.S. Patent No. 5,206,152 and WO 92/11033). Other  
conventional methods for gene delivery that can be used for delivery of the coding  
10 sequence include, for example, use of hand-held gene transfer particle gun (see, e.g., U.S.  
Patent No. 5,149,655); use of ionizing radiation for activating transferred gene (see, e.g.,  
U.S. Patent No. 5,206,152 and WO 92/11033).

The present invention will now be illustrated by reference to the following  
examples which set forth particularly advantageous embodiments. However, it should be  
15 noted that these embodiments are illustrative and are not to be construed as restricting the  
invention in any way.

## EXAMPLES

### EXAMPLE 1

#### SOURCE OF BIOLOGICAL MATERIALS AND OVERVIEW OF NOVEL POLYNUCLEOTIDES EXPRESSED BY THE BIOLOGICAL MATERIALS

5

Cell lines and human normal and tumor tissue were used to construct cDNA libraries from mRNA isolated from the cells and tissues. Most sequences were about 275-300 nucleotides in length. The cells lines include Km12L4-A cell line, a high metastatic colon cancer cell line (Morika, W. A. K. et al., *Cancer Research* (1988) 48:6863). The  
10 Km12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B2 surgical specimen (Morikawa et al. *Cancer Res.* (1988) 48:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman et al. *Nucl. Acids. Res.* (1995) 23:4007; Bao-Ling et al. *Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) 21:3269). The KM12C  
15 and KM12C-derived cell lines (e.g., KM12L4, KM12L4-A, etc.) are well-recognized in the art as model cell lines for the study of colon cancer (see, e.g., Moriakawa et al., *supra*; Radinsky et al. *Clin. Cancer Res.* (1995) 1:19; Yeatman et al., (1995) *supra*; Yeatman et al., *Clin. Exp. Metastasis* (1996) 14:246). These and other cell lines and tissue are described in Table 6.

20 The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol.* 266:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis  
25 Techniques" Adams et al., eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie et al. *Comput. Chem.* (1993) 17:191 ). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats. The sequences remaining after masking  
30 were then used in a BLASTN vs. Genbank search; sequences that exhibited greater than 70% overlap, 99% identity, and a p value of less than  $1 \times 10^{-40}$  were discarded. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database search: (1) unknown (no hits in the Genbank search), (2) weak similarity (greater than 45% identity and p value of less than  $1 \times 10^{-5}$ ), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than  $1 \times 10^{-5}$ ). Sequences having greater than 70% overlap, greater than 99% identity, and p value of less than  $1 \times 10^{-40}$  were discarded.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search was performed and sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than  $1 \times 10^{-40}$  were discarded. Sequences with a p value of less than  $1 \times 10^{-65}$  when compared to a database sequence of human origin were also excluded. Second, a BLASTN vs. Patent GeneSeq database was performed and sequences having greater than 99% identity, p value less than  $1 \times 10^{-40}$ , and greater than 99% overlap were discarded.

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than  $1 \times 10^{-111}$  in relation to a database sequence of human origin were specifically excluded. The final result provided the 3351 sequences listed in the accompanying Sequence Listing. Each identified polynucleotide represents sequence from at least a partial mRNA transcript. Polynucleotides that were determined to be novel were assigned a sequence identification number.

The novel polynucleotides were assigned sequence identification numbers SEQ ID NOs:1-3351. The first 1847 DNA sequences corresponding to the novel polynucleotides are provided in the Sequence Listing in Table 1. DNA sequences corresponding to the novel polynucleotides of SEQ ID NOs:1848-3351 are provided in the Sequence Listing in Table 2. The DNA sequences of Table 2, while numbered SEQ ID 1-1504, correspond to SEQ ID NOs:1848-3351 in the Sequence Listing, *e.g.*, Table 2 SEQ ID 1 is SEQ ID NO:1848, Table 2 SEQ ID 2 is SEQ ID NO:1849, *etc.* Each DNA sequence in Table 4 is uniquely identified by a number that is 1847 less than its SEQ ID NO in the Sequence Listing. Tables 1 and 2 provide: 1) the SEQ ID NO assigned to each sequence for use in the present specification or a corresponding number; 2) the sequence name used as an internal identifier of the sequence; 3) the name assigned to the clone from which the sequence was isolated; and 4) the number of the cluster to which the sequence is assigned (Cluster ID; where the cluster ID is 0, the sequence was not assigned to any cluster).

Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOs are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene.

## EXAMPLE 2

### RESULTS OF PUBLIC DATABASE SEARCH TO IDENTIFY FUNCTION OF GENE PRODUCTS

SEQ ID NOs:1-3351 were translated in all three reading frames to determine the best alignment with the individual sequences. These amino acid sequences and nucleotide sequences are referred to, generally, as query sequences, which are aligned with the individual sequences. Query and individual sequences were aligned using the BLAST programs, available over the world wide web at <http://www.ncbi.nlm.nih.gov/BLAST/>. Again the sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above in Example 1.

Tables 3 and 4 (inserted before the claims) show the results of the alignments. Table 3 contains alignment information for SEQ ID NOs:1-1847 and Table 4 contains alignment information for SEQ ID NOs:1848-3351. The DNA sequences of Table 4, while numbered SEQ ID 1-1504, correspond to SEQ ID NOs:1848-3351. Each DNA sequence in Table 4 is uniquely identified by a number that is 1847 less than its SEQ ID NO. Tables 3 and 4 refer to each sequence by its SEQ ID NO or a corresponding number, the accession numbers and descriptions of nearest neighbors from the Genbank and Non-Redundant Protein searches, and the p values of the search results.

For each of SEQ ID NOs:1-1847, the best alignment to a protein or DNA sequence is included in Table 3, and the best alignment for each of SEQ ID NOs:1848-3351 is included in Table 4. The activity of the polypeptide encoded by SEQ ID NOs:1-3351 is the same or similar to the nearest neighbor reported in Table 3 or 4. The accession number of the nearest neighbor is reported, providing a reference to the activities exhibited by the nearest neighbor. The search program and database used for the alignment also are indicated as well as a calculation of the p value.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence



of SEQ ID NOs:1-3351. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of SEQ ID NOs:1-3351.

### EXAMPLE 3

#### MEMBERS OF PROTEIN FAMILIES

5

The sequences (SEQ ID NOs:1-3351) were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or

10 comprising a known functional domain (Table 5). "Start" and "stop" in Table 3 indicate the position within the individual sequences that align with the query sequence having the indicated SEQ ID NO. The direction indicates the orientation of the query sequence with respect to the individual sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse

15 (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Some polynucleotides exhibited multiple profile hits because, for example, the particular sequence contains overlapping profile regions, and/or the sequence contains two different functional domains. These profile hits are described in more detail below.

20 Ank Repeats (ANK). SEQ ID NOs:187, 1268, 1804, 1819, 1830, 1839, 2652, 3015 and 3267 represent polynucleotides encoding an Ank repeat-containing protein. The ankyrin motif is a 33 amino acid sequence named for the protein ankyrin which has 24 tandem 33-amino-acid motifs. Ank repeats were originally identified in the cell-cycle-control protein cdc10 (Breedon et al., *Nature* (1987) 329:651). Proteins containing ankyrin

25 repeats include ankyrin, myotropin, I-kappaB proteins, cell cycle protein cdc10, the Notch receptor (Matsuno et al., *Development* (1997) 124(21):4265); G9a (or BAT8) of the class III region of the major histocompatibility complex (*Biochem J.* 290:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, and SW16. The functions of the ankyrin repeats are compatible with a role in protein-protein interactions (Bork, *Proteins* (1993) 17(4):363;

30 Lambert and Bennet, *Eur. J. Biochem.* (1993) 211:1; Kerr et al., *Current Op. Cell Biol.* (1992) 4:496; Bennet et al., *J. Biol. Chem.* (1980) 255:6424).

ATPases Associated with Various Cellular Activities (ATPases). Sequences within SEQ ID NOs:431, 639, 2135, 2684, 2859, 3197 and 3266 correspond to a sequence that encodes a novel member of the "ATPases Associated with diverse cellular Activities"

(AAA) protein family. The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids that contains an ATP-binding site (Froehlich et al., *J. Cell Biol.* (1991) 114:443; Erdmann et al., *Cell* (1991) 64:499; Peters et al., *EMBO J.* (1990) 9:1757; Kunau et al., *Biochimie* (1993) 75:209-224; 5 Confalonieri et al., *BioEssays* (1995) 17:639; <http://yeamob.pci.chemie.uni-tuebingen.de/AAA/Description.html>). The proteins that belong to this family either contain one or two AAA domains. In general, the AAA domains in these proteins act as ATP-dependent protein clamps (Confalonieri et al. (1995) *BioEssays* 17:639). In addition to the ATP-binding 'A' and 'B' motifs, which are located in the N-terminal half of this domain, 10 there is a highly conserved region located in the central part of the domain which was used in the development of the signature pattern. The consensus pattern is: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]-D-x-A-[LIFA]-x-R.

Bromodomain (bromodomain). SEQ ID NO:1814 represents a polynucleotide encoding a polypeptide having a bromodomain region (Haynes et al., 1992, *Nucleic Acids* 15 *Res.* 20:2693-2603, Tamkun et al., 1992, *Cell* 68:561-572, and Tamkun, 1995, *Curr. Opin. Genet. Dev.* 5:473-477), which is a conserved region of about 70 amino acids. The bromodomain is thought to be involved in protein-protein interactions and may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation. The consensus pattern, which spans a major part of the 20 bromodomain, is: [STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENQTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-[LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY].

Basic Region Plus Leucine Zipper Transcription Factors (BZIP). SEQ ID NOs:410, 552, 768, 822, 836, 1288, 1365, 1454, 1540, 1549, 1556, 1557, 1563, 1622, 1630, 1704, 25 1808, 2363, 2424, 3147, 3152, 3158 and 3208 represent polynucleotides encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) 2:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) 4:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine 30 zipper required for dimerization. The consensus pattern for this protein family is: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK].

EF Hand (EFhand). SEQ ID NOs:820, 1755 and 3285 correspond to polynucleotides encoding a novel protein in the family of EF-hand proteins. Many calcium-binding proteins belong to the same evolutionary family and share a type of 35 calcium-binding domain known as the EF-hand (Kawasaki et al., *Protein. Prof.* (1995)

2:305-490). This type of domain consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand). The consensus pattern includes the complete EF-hand loop as well as the first residue which follows the loop and which seem to always be hydrophobic: D-x-[DNS]-{ILVFIYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFIYW].

- 10        Ets Domain (Ets Nterm). SEQ ID NO:1811 represents a polynucleotide encoding a polypeptide with N-terminal homology in ETS domain. Proteins of this family contain a conserved domain, the "ETS-domain," that is involved in DNA binding. The domain appears to recognize purine-rich sequences; it is about 85 to 90 amino acids in length, and is rich in aromatic and positively charged residues (Wasylyk, et al., *Eur. J. Biochem.* (1993) 211:718). The *ets* gene family encodes a novel class of DNA-binding proteins, each of which binds a specific DNA sequence and comprises an *ets* domain that specifically interacts with sequences containing the common core tri-nucleotide sequence GGA. In addition to an *ets* domain, native *ets* proteins comprise other sequences which can modulate the biological specificity of the protein. *Ets* genes and proteins are involved in a variety of essential biological processes including cell growth, differentiation and development, and three members are implicated in oncogenic process.

- 25        G-Protein Alpha Subunit (G-alpha). SEQ ID NO:1846 represents a polynucleotide encoding a novel polypeptide of the G-protein alpha subunit family. Guanine nucleotide binding proteins (G-proteins) are a family of membrane-associated proteins that couple extracellularly-activated integral-membrane receptors to intracellular effectors, such as ion channels and enzymes that vary the concentration of second messenger molecules. G-proteins are composed of 3 subunits (alpha, beta and gamma) which, in the resting state, associate as a trimer at the inner face of the plasma membrane. The alpha subunit binds GTP and exhibits GTPase activity. G-protein alpha subunits are 350-400 amino acids in length and have molecular weights in the range 40-45 kDa. Seventeen distinct types of alpha subunit have been identified in mammals, and fall into 4 main groups on the basis of both sequence similarity and function: alpha-s, alpha-q, alpha-i and alpha-12 (Simon et al., *Science* (1993) 252:802). They are often N-terminally acylated, usually with myristate and/or palmitoylate, and these fatty acid modifications can be important for membrane association and high-affinity interactions with other proteins.

Helicases conserved C-terminal domain (helicase\_C). SEQ ID NOs:1496, 2826 and 2871 represent polynucleotides encoding novel members of the DEAD/H helicase family. A number of eukaryotic and prokaryotic proteins have been characterized (Schmid S.R., et al., *Mol. Microbiol.* (1992) 6:283; Linder P., et al., *Nature* (1989) 337:121; Wassarman D.A., et al., *Nature* (1991) 349:463) on the basis of their structural similarity. All are involved in ATP-dependent, nucleic-acid unwinding. All DEAD box family members of the above proteins share a number of conserved sequence motifs, some of which are specific to the DEAD family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman T.C., *Nature* (1988) 333:22 and *Nature* (1988) 333:578 (Errata). One of these motifs, called the "D-E-A-D-box", represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be "D-E-A-H-box" proteins (Wassarman D.A., et al., *Nature* (1991) 349:463; Harosh I., et al., *Nucleic Acids Res.* (1991) 19:6331; Koonin E.V. et al., *J. Gen. Virol.* (1992) 73:989. The following signature patterns are used to identify members of both subfamilies: 1) [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN]; and 2) [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR].

Homeobox domain (homeobox). SEQ ID NOs:1676, 1820 and 1821 represent polynucleotides encoding proteins having a homeobox domain. The homeobox is a protein domain of 60 amino acids (Gehring In: Guidebook to the Homeobox Genes, Duboule D., Ed., pp. 1-10, Oxford University Press, Oxford, (1994); Buerklin In: Guidebook to the Homeobox Genes, pp25-72, Oxford University Press, Oxford, (1994); Gehring, *Trends Biochem. Sci.* (1992) 17:277-280; Gehring et al., *Annu. Rev. Genet.* (1986) 20:147-173; Schofield, *Trends Neurosci.* (1987) 10:3-6) first identified in a number of *Drosophila* homeotic and segmentation proteins. It is extremely well conserved in many other animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Several proteins that contain a homeobox domain play an important role in development. Most of these proteins are sequence-specific DNA-binding transcription factors. The homeobox domain is also very similar to a region of the yeast mating type proteins. These are sequence-specific DNA-binding proteins that act as master switches in yeast differentiation by controlling gene expression in a cell type-specific fashion.

A schematic representation of the homeobox domain is shown below. The helix-turn-helix region is shown by the symbols 'H' (for helix), and 't' (for turn).

XX

1

60

The pattern detects homeobox sequences 24 residues long and spans positions 34 to  
 5 57 of the homeobox domain. The consensus pattern is as follows: [LIVMFYVG]-  
 [ASLVR]-x(2)-[LIVMSTACN]-x-[LIVM]-x(4)-[LIV]-[RKNQESTAIY]-[LIVFSTNKH]-  
 W-[FYVC]-x-[NDQTAH]-x(5)-[RKNAIMW].

MAP kinase kinase (mkk). SEQ ID NOs:29, 31, 196, 3175, 3190 and 3281  
 represent novel members of the MAP kinase kinase family. MAP kinases (MAPK) are  
 10 involved in signal transduction, and are important in cell cycle and cell growth controls.  
 The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which  
 phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast,  
 invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation  
 switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates.  
 15 MAPKKs are essential transducers through which signals must pass before reaching the  
 nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) 79:193-207; Nishida et al.,  
*Trends Biochem Sci* (1993) 18:128-31; Ruderman, *Curr Opin Cell Biol* (1993) 5:207-13;  
 Dhanasekaran et al., *Oncogene* (1998) 17:1447-55; Kiefer et al., *Biochem Soc Trans* (1997)  
 25:491-8; and Hill, *Cell Signal* (1996) 8:533-44.

Protein Kinase (protkinase). SEQ ID NOs:1157, 1478, 1496, 2286, 2969 and 3190  
 represent polynucleotides encoding protein kinases. Protein kinases catalyze  
 phosphorylation of proteins in a variety of pathways, and are implicated in cancer.  
 Eukaryotic protein kinases (Hanks S.K., et al., *FASEB J.* (1995) 9:576; Hunter T., *Meth.*  
*Enzymol.* (1991) 200:3; Hanks S.K., et al., *Meth. Enzymol.* (1991) 200:38; Hanks S.K.,  
 25 *Curr. Opin. Struct. Biol.* (1991) 1:369; Hanks S.K. et al., *Science* (1988) 241:42) are  
 enzymes that belong to a very extensive family of proteins which share a conserved  
 catalytic core common to both serine/threonine and tyrosine protein kinases. There are a  
 number of conserved regions in the catalytic domain of protein kinases. The first region,  
 which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich  
 30 stretch of residues in the vicinity of a lysine residue, which has been shown to be involved  
 in ATP binding. The second region, which is located in the central part of the catalytic  
 domain, contains a conserved aspartic acid residue which is important for the catalytic  
 activity of the enzyme (Knighton D.R. et al., *Science* (1991) 253:407). The protein kinase  
 profile includes two signature patterns for this second region: one specific for  
 35 serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the

alignment in (Hanks S.K. et al., *FASEB J.* (1995) 9:576) and covers the entire catalytic domain.

The consensus patterns are as follows: 1) [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-

- 5 [AIVP]-[LIVMFAGCKR]-K, where K binds ATP; 2) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3), where D is an active site residue; and 3) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC], where D is an active site residue.

- 10 If a protein analyzed includes two of the above protein kinase signatures, the probability of it being a protein kinase is close to 100%.

- Ras family proteins (ras). SEQ ID NOs:1688 and 3258 represent polynucleotides encoding novel members of the ras family of small GTP/GDP-binding proteins (Valencia et al., 1991, *Biochemistry* 30:4637-4648). Ras family members generally require a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. Among ras-related proteins, the highest degree of sequence conservation is found in four regions that are directly involved in guanine nucleotide binding. The first two constitute most of the phosphate and Mg<sup>2+</sup> binding site (PM site) and are located in the first half of the G-domain. The other two regions are involved in guanosine binding and are located in the C-terminal half of the molecule.
- 20 Motifs and conserved structural features of the ras-related proteins are described in Valencia et al., 1991, *Biochemistry* 30:4637-4648. A major consensus pattern of ras proteins is: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y.

- Thioredoxin family active site (Thioredoxin). SEQ ID NO:1677 represents a polynucleotide encoding a protein having a thioredoxin family active site. Thioredoxins (Holmgren A., *Annu. Rev. Biochem.* (1985) 54:237; Gleason F.K. et al., *FEMS Microbiol. Rev.* (1988) 54:271; Holmgren, A. *J. Biol. Chem.* (1989) 264:13963; Eklund H. et al., *Proteins* (1991) 11:13) are small proteins of approximately one hundred amino- acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is well conserved. All PDI contains two or three (ERp72) copies of the thioredoxin domain. The consensus pattern is: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGNT]-C- [GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT] (where the two C's
- 35 form the redox-active bond).

Trypsin (trypsin). SEQ ID NO:1410 corresponds to a novel serine protease of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature* (1988) 334:528). The consensus patterns for this trypsin protein family are: 1) [LIVM]-[ST]-A-[STAG]-H-C, where H is the active site residue; and 2) [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-[LIVMFYSTANQH], where S is the active site residue. All sequences known to belong to this family are detected by the above consensus sequences, except for 18 different proteases which have lost the first conserved glycine. If a protein includes both the serine and the histidine active site signatures, the probability of it being a trypsin family serine protease is 100%.

WD Domain, G-Beta Repeats (WD domain). SEQ ID NOs:1336, 1380, 1711, 1762, 1909, 2218, 3047, 3108 and 3292 represent novel members of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) 56:615). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat). The consensus pattern for the WD domain/G-Beta repeat family is: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-[LIVMWSTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN].

wnt Family of Developmental Signaling Proteins (Wnt dev sign). SEQ ID NO:1538 corresponds to a novel member of the wnt family of developmental signaling proteins. Wnt-1 (previously known as int-1), the seminal member of this family, (Nusse R., *Trends Genet.* (1988) 4:291) is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous system (CNS). All wnt family proteins share the following features characteristics of secretory proteins: a signal peptide, several potential N-glycosylation sites and 22

conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters. The consensus pattern, which is based upon a highly conserved region including three cysteines, is as follows: C-K-C-H-G-[LIVMT]-S-G-x-C.

- 5        Protein Tyrosine Phosphatase (Y\_phosphatase). SEQ ID NO:1417 represents a polynucleotide encoding a protein tyrosine kinase. Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) (Fischer et al., *Science* (1991) 253:401; Charbonneau et al., *Annu. Rev. Cell Biol.* (1992) 8:463; Trowbridge, *J. Biol. Chem.* (1991) 266:23517; Tonks et al., *Trends Biochem. Sci.* (1989) 14:497; and Hunter, *Cell* (1989) 58:1013) catalyze the
- 10 removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). Structurally, all known receptor PTPases are made up of a variable length
- 15 extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. PTPase domains consist of about 300 amino acids. The search of two conserved cysteines has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important. The consensus pattern for PTPases is: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-
- 20 [STAGP]-x-[LIVMFY]; C is the active site residue.

- Zinc Finger, C2H2 Type (Zincfing\_C2H2). SEQ ID NOs:308, 807, 1324, 1503, 1527, 3081, 3193 and 3306 correspond to polynucleotides encoding novel members of the of the C2H2 type zinc finger protein family. Zinc finger domains (Klug et al., *Trends Biochem. Sci.* (1987) 12:464; Evans et al., *Cell* (1988) 52:1; Payre et al., *FEBS Lett.* (1988)
- 25 234:245; Miller et al., *EMBO J.* (1985) 4:1609; and Berg, *Proc. Natl. Acad. Sci. USA* (1988) 85:99) are nucleic acid-binding protein structures. In addition to the conserved zinc ligand residues, it has been shown that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld et al., *J. Biomol. Struct. Dyn.* (1993) 11:557) The best conserved position is found four residues after the second
- 30 cysteine; it is generally an aromatic or aliphatic residue. The consensus pattern for C2H2 zinc fingers is: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H. The two C's and two H's are zinc ligands.

- Src homology 2. SEQ ID NOs:186, 2591, 3307 and 3339 represent polynucleotides encoding novel members of the family of Src homology 2 (SH2) proteins. The Src
- 35 homology 2 (SH2) domain is a protein domain of about 100 amino acid residues first



identified as a conserved sequence region between the oncoproteins Src and Fps (Sadowski I. et al., *Mol. Cell. Biol.* 6:4396-4408 (1986)). Similar sequences are found in many other intracellular signal-transducing proteins (Russel R.B. et al., *FEBS Lett.* 304:15-20 (1992)). SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and phosphorylation-dependent manner (Marangere L.E.M., Pawson T., *J. Cell Sci. Suppl.* 18:97-104 (1994); Pawson T., Schlessinger J., *Curr. Biol.* 3:434-442 (1993); Mayer B.J., Baltimore D., *Trends Cell. Biol.* 3:8-13 (1993); Pawson T., *Nature* 373:573-580 (1995)).

10       The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets (Kuriyan J., Cowburn D., *Curr. Opin. Struct. Biol.* 3:828-837(1993)). The profile to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions.

15       Src homology 3. SEQ ID NO:234, 1832, and 1835 represent polynucleotides encoding novel members of the family of Src homology 3 (SH3) proteins. The Src homology 3 (SH3) domain is a small protein domain of about 60 amino acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (e.g., Src, Abl, Lck) (Mayer B.J. et al., *Nature* 332:272-275 (1988)). Since  
20 then, it has been found in a great variety of other intracellular or membrane-associated proteins (Musacchio A. et al., *FEBS Lett.* 307:55-61 (1992); Pawson T., Schlessinger J., *Curr. Biol.* 3:434-442 (1993); Mayer B.J., Baltimore D., *Trends Cell Biol.* 3:8-13 (1993); Pawson T., *Nature* 373:573-580 (1995)).

25       The SH3 domain has a characteristic fold which consists of five or six beta strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices (Kuriyan J., Cowburn D., *Curr. Opin. Struct. Biol.* 3:828-837 (1993)).

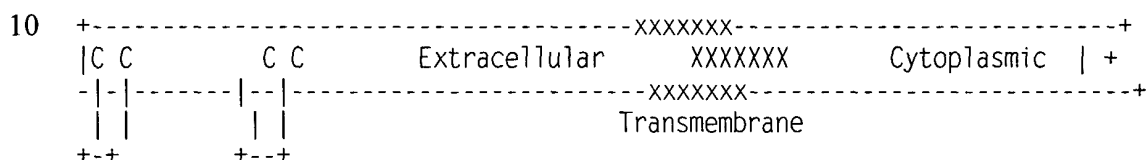
      The function of the SH3 domain may be to mediate assembly of specific protein complexes via binding to proline-rich peptides (Morton C.J., Campbell I.D., *Curr. Biol.* 4:615-617 (1994)).

30       In general SH3 domains are found as single copies in a given protein, but there are a significant number of proteins with two SH3 domains and a few with 3 or 4 copies.

Fibronectin type III. SEQ ID NOs:746 and 1192 represent polynucleotides encoding novel members of the family of fibronectin type III proteins. A number of receptors for lymphokines, hematopoietic growth factors and growth hormone-related  
35 molecules have been found to share a common binding domain. (Bazan J.F., *Biochem.*

*Biophys. Res. Commun.* 164:788-795 (1989); Bazan J.F., *Proc. Natl. Acad. Sci. U.S.A.* 87:6934-6938 (1990); Cosman D. et al., *Trends Biochem. Sci.* 15:265-270 (1990); d'Andrea A.D., Fasman G.D., Lodish H.F., *Cell* 58:1023-1024 (1989); d'Andrea A.D., Fasman G.D., Lodish H.F., *Curr. Opin. Cell Biol.* 2:648-651 (1990)).

- 5 The conserved region constitutes all or part of the extracellular ligand- binding region and is about 200 amino acid residues long. In the N-terminal of this domain there are two pairs of cysteines known, in the growth hormone receptor, to be involved in disulfide bonds.



Two patterns detect this family of receptors. The first one is derived from the first N-terminal disulfide loop, the second is a tryptophan-rich pattern located at the C-terminal extremity of the extracellular region.

- A consensus for this protein family is: C-[LVFYR]-x(7,8)-[STIVDN]-C-x-W (The two C's are linked by a disulfide bond]. A second consensus for this protein family is: [STGL]-x-W-[SG]-x-W-S.
- 20

- LIM domain containing proteins. SEQ ID NOs:1269, 1309, 1360, and 1386 represent polynucleotides encoding novel members of the family of LIM domain containing proteins. A number of proteins contain a conserved cysteine-rich domain of about 60 amino-acid residues. (Freyd G. et al., *Nature* 344:876-879 (1990); Baltz R. et al., *Plant Cell* 4:1465-1466 (1992); Sanchez-Garcia I., Rabbitts T.H., *Trends Genet.* 10:315-320 (1994)).
- 25

- In the LIM domain, there are seven conserved cysteine residues and a histidine. The arrangement followed by these conserved residues is C-x(2)-C-x(16,23)-H-x(2)-[CH]-x(2)-C-x(2)-C-x(16,21)-C-x(2,3)-[CHD]. The LIM domain binds two zinc ions (Michelsen J.W. et al., *Proc. Natl. Acad. Sci. U.S.A.* 90:4404-4408 (1993)). LIM does not bind DNA, rather it seems to act as interface for protein-protein interaction. The consensus for this protein family is: C-x(2)-C-x(15,21)-[FYWH]-H-x(2)-[CH]-x(2)-C-x(2)-C-x(3)-[LIVMF]. The 5 C's and the H bind zinc.
- 30

- C2 domain (protein kinase C like). SEQ ID NOs:1325 and 2282 represent polynucleotides encoding novel members of the family of C2 domain containing proteins. Some isozymes of protein kinase C (PKC) contain a domain, known as C2, of about 116
- 35

amino-acid residues, which is located between the two copies of the C1 domain (that bind phorbol esters and diacylglycerol) and the protein kinase catalytic domain. (Azzi A. et al., *Eur. J. Biochem.* 208:547-557 (1992); Stabel S., *Semin. Cancer Biol.* 5:277-284 (1994)).

The C2 domain is involved in calcium-dependent phospholipid binding (Davletov B.A., Suedhof T.C., *J. Biol. Chem.* 268:26386-26390 (1993)). Since domains related to the C2 domain are also found in proteins that do not bind calcium, other putative functions for the C2 domain include binding to inositol-1,3,5-tetraphosphate. (Fukuda M., et al., *J. Biol. Chem.* 269:29206-29211 (1994).)

The consensus pattern for the C2 domain is located in a conserved part of that domain, the connecting loop between beta strands 2 and 3. The profile for the C2 domain covers the total domain. The consensus for this protein family is:: [ACG]-x(2)-L-x(2,3)-D-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-D- [PA]-[FY]

Serine proteases, trypsin family, active sites. SEQ ID NO:1410 represents a polynucleotide encoding a novel member of the family of serine protease, trypsin proteins. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature* 334:528-530 (1988)).

A consensus for this protein family is: [LIVM]-[ST]-A-[STAG]-H-C [H is the active site residue]. A second consensus for this protein family is: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]- [LIVMFYWH]-[LIVMFYSTANQH] [S is the active site residue].

RNA Recognition Motif Domain (RRM, RBD, or RNP). SEQ ID NOS: 1464 and 1514 represent polynucleotides encoding novel members of the family of RNA recognition motif domain proteins (Bandziulis R.J. et al., *Genes Dev.* 3:431-437 (1989); Dreyfuss G. et al., *Trends Biochem. Sci.* 13:86-91 (1988)).

Inside the putative RNA-binding domain there are two regions which are highly conserved. The first one is a hydrophobic segment of six residues (which is called the RNP-2 motif); the second one is an octapeptide motif (which is called RNP-1 or RNP-CS). The position of both motifs in the domain is shown in the following schematic representation:

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
          RNP-2                                RNP-1
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

```

As a consensus pattern for this type of domain the RNP-1 motif was used. The consensus for this protein family is: [RK]-G-{EDRKHPCG}-[AGSCI]-[FY]-[LIVA]-x-[FYLM]

Phosphatidylinositol-specific phospholipase C, Y Domain. SEQ ID NO: 1707  
5 represents a polynucleotide encoding a novel member of the phosphatidylinositol-specific phospholipase C, Y domain family of proteins. Phosphatidylinositol-specific phospholipase C (EC3.1.4.11), a eukaryotic intracellular enzyme, plays an important role in signal transduction processes (Meldrum E. et al., *Biochim. Biophys. Acta* 1092:49-71 (1991)). It catalyzes the hydrolysis of 1-phosphatidyl-D-myo-inositol-3,4,5- triphosphate  
10 into the second messenger molecules diacylglycerol and inositol-1,4,5-triphosphate. This catalytic process is tightly regulated by reversible phosphorylation and binding of regulatory proteins (Rhee S.G., Choi K.D., *Adv. Second Messenger Phosphoprotein Res.* 26:35-61 (1992); Rhee S.G., Choi K.D., *J. Biol. Chem.* 267:12393-12396 (1992); Sternweis P.C., Smrcka A.V., *Trends Biochem. Sci.* 17:502-506 (1992)).

15 All eukaryotic PI-PLCs contain two regions of homology, referred to as "X-box" and "Y-box". The order of these two regions is the same (NH<sub>2</sub>-X-Y-COOH), but the spacing is variable. In most isoforms, the distance between these two regions is only 50-100 residues but in the gamma isoforms one PH domain, two SH2 domains, and one SH3 domain are inserted between the two PLC-specific domains. The two conserved regions  
20 have been shown to be important for the catalytic activity. At the C-terminal of the Y-box, there is a C2 domain possibly involved in Ca-dependent membrane attachment.

Serine Carboxypeptidases. SEQ ID NO:1744 represents a polynucleotide encoding a novel member of the serine carboxypeptidases family of proteins. Carboxypeptidases may be either metallo carboxypeptidases or serine carboxypeptidases (EC 3.4.16.5 and EC  
25 3.4.16.6). The catalytic activity of the serine carboxypeptidases, like that of the trypsin family serine proteases, is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which is itself hydrogen-bonded to a serine (Liao D.I., Remington S.J., *J. Biol. Chem.* 265:6528-6531 (1990)).

The sequences surrounding the active site serine and histidine residues are highly  
30 conserved in all these serine carboxypeptidases. A consensus for this protein family is: [LIVM]-x-[GTA]-E-S-Y-[AG]-[GS] [S is the active site residue]. A second consensus for this protein family is: [LIVF]-x(2)-[LIVSTA]-x-[IVPST]-x-[GSDNQL]-[SAGV]-[SG]-H-x- [IVAQ]-P-x(3)-[PSA] [H is the active site residue].

dsrcm Double-Stranded RNA Binding Motif. SEQ ID NO:1818 represents a  
35 polynucleotide encoding a novel member of the dsrcm double-stranded RNA binding motif

proteins. In eukaryotic cells, a multitude of RNA-binding proteins play key roles in the posttranscriptional regulation of gene expression. Characterization of these proteins has led to the identification of several RNA-binding motifs. Several human and other vertebrate genetic disorders are caused by aberrant expression of RNA-binding proteins. (C. G. Burd  
 5 & G. Dreyfuss, *Science* 265: 615-621 (1994)).

Proteins containing double stranded RNA binding motifs bind to specific RNA targets. Double stranded RNA binding motifs are exemplified by interferon-induced protein kinase in humans, which is part of the cellular response to dsRNA.

SEQ ID NOs:2577, 3183 and 3195 encode members of the 4 trans-membrane  
 10 integral membrane protein family. This family consists of type III proteins, which are integral membrane proteins that contain a N-terminal membrane-anchoring domain that is not cleaved during biosynthesis, and which functions as a translocation signal and a membrane anchor. The proteins also have three additional transmembrane regions. The consensus pattern is: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF] (2)-G-C-x-[GA]-[STA]-x(20-  
 15 [eG]-x(20-[CwN]-[LIVM](2).

SEQ ID NO:2944 encodes a polypeptide having a calpain large subunit, domain III. Calpains are a family of intracellular proteases that play a variety of biological roles. Calpain 3, also known as p94, is predominantly expressed in skeletal muscle and plays a role in limb-girdle muscular dystrophy type 2A. (Sorimachi, H. et al., *Biochem. J.*  
 20 328:721-732, 1997).

SEQ ID NOs:1911 and 1980 encode polypeptides having a C3HC4 type zinc finger domain (RING finger), which is a cysteine-rich domain of 40 to 60 residues that binds two atoms of zinc, and is believed to be involved in mediating protein-protein interactions. Mammalian proteins of this family include V(D)J recombination activating protein, which  
 25 activates the rearrangement of immunoglobulin and T-cell receptor genes; breast cancer type 1 susceptibility protein (BRCA1); bmi-1 proto-oncogene; cbl proto-oncogene; and mel-18 protein, which is expressed in a variety of tumor cells and is a transcriptional repressor that recognizes and binds a specific DNA sequence. The consensus pattern is: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA].

30 SEQ ID NO:3274 encodes a eukaryotic transcription factor with a fork head domain, of about 100 amino acid residues. Proteins of this group are transcription factors, including mammalian transcription factors HNF-3-alpha, -beta, and -gamma; interleukin-enhancer binding factor; and HTLF, which binds to a region of human T-cell leukemia virus long terminal repeat. The consensus pattern is [KR]-P-[PTQ]-[FYLVQH]-S-  
 35 [FY]x(2)-[LIVM]-X(3,4)-[AC]-[LIM].

SEQ ID NO:3345 encodes a polypeptide having a PDZ domain. Several dozen signaling proteins belong to this group of proteins that have 80-100 residue repeats known as PDZ domains. Several of the proteins interact with the C-terminal tetrapeptide motifs X-Ser/Thr/X-Val-COO- of ion channels and/or receptors. (Ponting, C. P., Protein Sci. 6;464-468, 1997.)

SEQ ID NO:3351 encodes a polypeptide in the family of phorbol esters/glycerol binding proteins. Phorbol esters (PE) are analogues of diacylglycerol (DAG) and potent tumor promoters. DAG activates a family of serine-threonine protein kinases, known as protein kinase C. The N-terminal region of protein kinase C binds PE and DAG, and contains one or two copies of a cysteine-rich domain of about 50 amino acid residues. Other proteins having this domain include diacylglycerol kinase; the vav oncogene; and N-chimaerin, a brain-specific protein. The DAG/PE binding domain binds two zinc ions through the six cysteines and two histidines that are conserved in the domain. The consensus pattern is: H-x-[LIVMFYW]-x(8, 11)-C-x(2)-C-x(3)-[LIVMFC]-x(5, 10)-C-x(2)-C-x(4)-[HD]-x(2)-C-x(5, 9)-C.

SEQ ID NO:2216 encodes a polypeptide having a WW/rsp5/WWP domain. The protein is named for the presence of conserved aromatic positions, generally tryptophan, as well as a conserved proline. Proteins having the domain include dystrophin, vertebrate YAP protein, and IQGAP, a human GTPase activating protein which acts on ras. The consensus pattern is: W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

SEQ ID NO:2428 encodes a member of the dual specificity phosphatase family, having a catalytic domain, and SEQ IDS NOs:2281 and 2310 encode members of the protein tyrosine phosphatase family. These families are related and classified as tyrosine specific protein phosphatases. The enzymes catalyze the removal of a phosphate group from a tyrosine residue, and are important in the control of cell growth, proliferation, differentiation, and transformation. The consensus pattern is [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY].

Table 5

| SEQ ID | Start | Stop | Score | Direction | Description  |
|--------|-------|------|-------|-----------|--|
| 29     | 295   | 421  | 5872  | For       | mkk like kinases                                       |
| 30     | 31    | 182  | 3943  | For       | Basic region plus leucine zipper transcription factors |
| 31     | 298   | 397  | 5625  | For       | mkk like kinases                                       |
| 186    | 175   | 395  | 7660  | For       | SH2 Domain   |
| 187    | 358   | 432  | 4320  | For       | Ank repeat   |
| 196    | 37    | 322  | 6049  | For       | mkk like kinases                                       |
| 234    | 23    | 121  | 4607  | For       | SH3 Domain   |
| 308    | 110   | 172  | 4150  | For       | Zinc finger, C2H2 type                                 |
| 410    | 42    | 191  | 4036  | For       | Basic region plus leucine zipper transcription factors |
| 431    | 71    | 428  | 5538  | Rev       | ATPases Associated with Various Cellular Activities    |
| 552    | 116   | 288  | 3930  | Rev       | Basic region plus leucine zipper transcription factors |
| 639    | 157   | 561  | 5797  | For       | ATPases Associated with Various Cellular Activities    |
| 746    | 209   | 427  | 5379  | For       | Fibronectin type III domain                            |
| 768    | 116   | 288  | 3930  | For       | Basic region plus leucine zipper transcription factors |
| 807    | 339   | 392  | 3620  | For       | Zinc finger, C2H2 type                                 |
| 820    | 341   | 406  | 2930  | Rev       | EF-hand  |
| 822    | 108   | 262  | 4179  | For       | Basic region plus leucine zipper transcription factors |
| 836    | 158   | 353  | 4430  | For       | Basic region plus leucine zipper transcription factors |
| 1157   | 41    | 444  | 5279  | Rev       | protein kinase   |
| 1192   | 186   | 416  | 5469  | For       | Fibronectin type III domain                            |
| 1268   | 238   | 315  | 3540  | For       | Ank repeat   |
| 1269   | 79    | 240  | 11640 | For       | LIM domain containing proteins                         |
| 1288   | 73    | 234  | 3953  | For       | Basic region plus leucine zipper transcription factors |
| 1309   | 248   | 404  | 8226  | for       | LIM domain containing proteins                         |

| SEQ ID | Start | Stop | Score | Direction | Description  |
|--------|-------|------|-------|-----------|--|
| 1324   | 294   | 356  | 4690  | for       | Zinc finger, C2H2 type                                 |
| 1325   | 1     | 234  | 8981  | for       | C2 domain (prot. kinase C like)                        |
| 1336   | 66    | 164  | 6390  | for       | WD domain, G-beta repeats                              |
| 1360   | 222   | 377  | 8686  | for       | LIM domain containing proteins                         |
| 1365   | 69    | 257  | 5221  | for       | Basic region plus leucine zipper transcription factors |
| 1380   | 42    | 140  | 7130  | for       | WD domain, G-beta repeats                              |
| 1386   | 243   | 398  | 8736  | for       | LIM domain containing proteins                         |
| 1410   | 222   | 350  | 10553 | for       | Trypsin  |
| 1417   | 8     | 354  | 6073  | for       | Protein Tyrosine Phosphatase                           |
| 1454   | 49    | 209  | 3996  | for       | Basic region plus leucine zipper transcription factors |
| 1464   | 4     | 180  | 4978  | for       | RNA recognition motif. (aka RRM, RBD, or RNP domain)   |
| 1478   | 54    | 437  | 5176  | for       | protein kinase   |
| 1496   | 241   | 520  | 3929  | for       | Helicases conserved C-terminal domain                  |
| 1496   | 40    | 612  | 5187  | for       | protein kinase   |
| 1503   | 154   | 216  | 4870  | for       | Zinc finger, C2H2 type                                 |
| 1514   | 2     | 252  | 4662  | for       | RNA recognition motif. (aka RRM, RBD, or RNP domain)   |
| 1527   | 156   | 212  | 3520  | for       | Zinc finger, C2H2 type                                 |
| 1538   | 9     | 635  | 11087 | for       | wnt family of developmental signaling proteins         |
| 1540   | 289   | 471  | 4107  | for       | Basic region plus leucine zipper transcription factors |
| 1549   | 200   | 391  | 4118  | for       | Basic region plus leucine zipper transcription factors |
| 1556   | 163   | 354  | 3958  | for       | Basic region plus leucine zipper transcription factors |
| 1557   | 207   | 398  | 4038  | for       | Basic region plus leucine zipper transcription factors |
| 1563   | 107   | 298  | 3978  | for       | Basic region plus leucine zipper transcription factors |
| 1622   | 180   | 365  | 4022  | for       | Basic region plus leucine zipper transcription factors |
| 1630   | 100   | 291  | 3998  | for       | Basic region plus leucine zipper transcription factors |



| SEQ ID | Start | Stop | Score | Direction | Description   |
|--------|-------|------|-------|-----------|---|
| 1674   | 196   | 258  | 4880  | for       | Zinc finger, C2H2 type  |
| 1676   | 9     | 86   | 6610  | for       | Homeobox Domain   |
| 1677   | 316   | 369  | 5780  | rev       | Thioredoxins  |
| 1688   | 109   | 410  | 17414 | for       | Ras family  |
| 1704   | 184   | 372  | 3977  | for       | Basic region plus leucine zipper transcription factors                          |
| 1707   | 92    | 439  | 24100 | rev       | Phosphatidylinositol-specific phospholipase C, Y domain                         |
| 1711   | 263   | 361  | 6400  | for       | WD domain, G-beta repeats   |
| 1744   | 238   | 433  | 10572 | rev       | Serine carboxypeptidases  |
| 1755   | 281   | 367  | 2580  | for       | EF-hand   |
| 1762   | 236   | 334  | 5880  | for       | WD domain, G-beta repeats   |
| 1779   | 64    | 126  | 4790  | for       | Zinc finger, C2H2 type  |
| 1801   | 295   | 351  | 4030  | for       | Zinc finger, C2H2 type  |
| 1804   | 301   | 378  | 3460  | for       | Ank repeat  |
| 1808   | 36    | 161  | 4170  | for       | Basic region plus leucine zipper transcription factors                          |
| 1811   | 184   | 315  | 8390  | for       | N-terminal homology in Ets domain   |
| 1814   | 127   | 294  | 10770 | for       | Bromodomain (conserved sequence found in human, Drosophila and yeast proteins.) |
| 1818   | 9     | 146  | 4741  | for       | Double-stranded RNA binding motif   |
| 1819   | 278   | 355  | 3460  | for       | Ank repeat  |
| 1820   | 123   | 299  | 12150 | for       | Homeobox Domain   |
| 1821   | 127   | 303  | 12180 | for       | Homeobox Domain   |
| 1830   | 184   | 267  | 4270  | for       | Ank repeat  |
| 1832   | 18    | 173  | 8987  | for       | SH3 Domain  |
| 1835   | 51    | 206  | 8987  | for       | SH3 Domain  |
| 1839   | 224   | 307  | 4270  | for       | Ank repeat  |
| 1846   | 12    | 398  | 36700 | for       | G-protein alpha subunit   |
| 1909   | 160   | 258  | 6370  | for       | WD domain, G-beta repeats   |
| 1911   | 35    | 151  | 9335  | for       | Zinc finger, C3HC4 type (RING finger)   |
| 1980   | 60    | 197  | 7917  | for       | Zinc finger, C3HC4 type (RING finger)   |
| 2065   | 253   | 306  | 5410  | for       | Zinc finger, CCHC class   |
| 2135   | 2     | 401  | 10596 | for       | ATPases Associated with Various Cellular Activities                             |

| SEQ ID | Start | Stop | Score | Direction | Description  |
|--------|-------|------|-------|-----------|--|
| 2216   | 90    | 179  | 5380  | for       | WW/rsp5/WWP domain containing proteins                 |
| 2218   | 127   | 225  | 5500  | for       | WD domain, G-beta repeats                              |
| 2281   | 20    | 387  | 6044  | for       | Protein Tyrosine Phosphatase                           |
| 2282   | 183   | 353  | 5136  | for       | C2 domain (prot. kinase C like)                        |
| 2286   | 12    | 382  | 5228  | for       | protein kinase   |
| 2310   | 20    | 371  | 5962  | for       | Protein Tyrosine Phosphatase                           |
| 2363   | 48    | 211  | 4132  | for       | Basic region plus leucine zipper transcription factors |
| 2424   | 43    | 194  | 3996  | for       | Basic region plus leucine zipper transcription factors |
| 2428   | 25    | 350  | 4675  | for       | Dual specificity phosphatase, catalytic domain         |
| 2562   | 18    | 101  | 4560  | for       | Ank repeat   |
| 2577   | 0     | 311  | 10295 | for       | 4 transmembrane segments integral membrane proteins    |
| 2591   | 60    | 165  | 4560  | for       | SH2 Domain   |
| 2684   | 9     | 461  | 5759  | for       | ATPases Associated with Various Cellular Activities    |
| 2826   | 116   | 400  | 16107 | for       | DEAD and DEAH box helicases                            |
| 2859   | 100   | 320  | 5550  | rev       | ATPases Associated with Various Cellular Activities    |
| 2871   | 198   | 392  | 9384  | for       | DEAD and DEAH box helicases                            |
| 2944   | 18    | 281  | 10480 | for       | Calpain large subunit, domain III                      |
| 2969   | 5     | 387  | 5976  | rev       | protein kinase   |
| 3015   | 131   | 214  | 3600  | for       | Ank repeat   |
| 3047   | 191   | 292  | 5295  | for       | WD domain, G-beta repeats                              |
| 3081   | 190   | 252  | 4360  | for       | Zinc finger, C2H2 type                                 |
| 3108   | 275   | 367  | 5791  | for       | WD domain, G-beta repeats                              |
| 3147   | 190   | 369  | 4022  | for       | Basic region plus leucine zipper transcription factors |
| 3152   | 129   | 320  | 3947  | for       | Basic region plus leucine zipper transcription factors |
| 3158   | 167   | 334  | 4180  | for       | Basic region plus leucine zipper transcription factors |
| 3175   | 14    | 164  | 5951  | for       | mkk like kinases                                       |
| 3175   | 8     | 112  | 5968  | for       | protein kinase   |
| 3178   | 45    | 386  | 19398 | for       | ATPases Associated with Various Cellular Activities    |
| 3183   | 14    | 215  | 9133  | for       | 4 transmembrane segments integral membrane proteins    |
| 3190   | 229   | 390  | 6089  | for       | mkk like kinases                                       |
| 3190   | 118   | 390  | 8063  | for       | protein kinase   |
| 3193   | 293   | 355  | 3570  | for       | Zinc finger, C2H2 type                                 |

| SEQ ID | Start | Stop | Score | Direction | Description  |
|--------|-------|------|-------|-----------|--|
| 3195   | 0     | 215  | 10146 | for       | 4 transmembrane segments integral membrane proteins    |
| 3197   | 281   | 343  | 4490  | for       | Zinc finger, C2H2 type                                 |
| 3208   | 34    | 256  | 4190  | for       | Basic region plus leucine zipper transcription factors |
| 3258   | 138   | 394  | 9877  | for       | Ras family   |
| 3266   | 8     | 139  | 9328  | for       | ATPases Associated with Various Cellular Activities    |
| 3267   | 97    | 180  | 3820  | for       | Ank repeat   |
| 3274   | 11    | 187  | 15442 | for       | Fork head domain, eukaryotic transcription factors     |
| 3281   | 15    | 182  | 9681  | for       | mkk like kinases                                       |
| 3285   | 16    | 102  | 4680  | for       | EF-hand  |
| 3292   | 208   | 300  | 5585  | for       | WD domain, G-beta repeats                              |
| 3297   | 7     | 153  | 6100  | for       | Helicases conserved C-terminal domain                  |
| 3306   | 161   | 223  | 4900  | for       | Zinc finger, C2H2 type                                 |
| 3307   | 43    | 321  | 8740  | for       | SH2 Domain   |
| 3339   | 94    | 342  | 14970 | for       | SH2 Domain   |
| 3345   | 65    | 271  | 12512 | for       | PDZ domain   |
| 3351   | 124   | 270  | 6068  | for       | Phorbol esters/diacylglycerol binding                  |

## Example 4

## DIFFERENTIAL EXPRESSION OF POLYNUCLEOTIDES OF THE INVENTION:

## DESCRIPTION OF LIBRARIES AND DETECTION OF DIFFERENTIAL EXPRESSION

- 5 The relative expression levels of the polynucleotides of the invention was assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 6 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepare the cDNA library, the abbreviated name of the library that is used in the tables below (in quotes), and the approximate number
- 10 of clones in the library.

Table 6  
Description of cDNA Libraries

| Library<br>(lib #) | Description  | Number of<br>Clones in<br>this<br>Clustering |
|--------------------|--|--|
| 1                  | Km12 L4<br>Human Colon Cell Line, High Metastatic Potential<br>(derived from Km12C)<br>"High Colon"                  | 307133                                       |
| 2                  | Km12C<br>Human Colon Cell Line, Low Metastatic Potential<br>"Low Colon"  | 284755                                       |
| 3                  | MDA-MB-231<br>Human Breast Cancer Cell Line, High Metastatic Potential;<br>micro-metastases in lung<br>"High Breast" | 326937                                       |
| 4                  | MCF7<br>Human Breast Cancer Cell, Non Metastatic<br>"Low Breast"   | 318979                                       |
| 8                  | MV-522<br>Human Lung Cancer Cell Line, High Metastatic Potential<br>"High Lung"                                      | 223620                                       |
| 9                  | UCP-3<br>Human Lung Cancer Cell Line, Low Metastatic Potential<br>"Low Lung"   | 312503                                       |
| 12                 | Human microvascular endothelial cells (HMEC) – Untreated<br>PCR (OligodT) cDNA library                               | 41938  |

| Library (lib #) | Description   | Number of Clones in this Clustering |
|-----------------|---|-------------------------------------|
| 13              | Human microvascular endothelial cells (HMEC) – Basic fibroblast growth factor (bFGF) treated PCR (OligodT) cDNA library     | 42100                               |
| 14              | Human microvascular endothelial cells (HMEC) – Vascular endothelial growth factor (VEGF) treated PCR (OligodT) cDNA library | 42825                               |
| 15              | Normal Colon – UC#2 Patient PCR (OligodT) cDNA library “Normal Colon Tumor Tissue”  | 34285                               |
| 16              | Colon Tumor – UC#2 Patient PCR (OligodT) cDNA library “Normal Colon Tumor Tissue”   | 35625                               |
| 17              | Liver Metastasis from Colon Tumor of UC#2 Patient PCR (OligodT) cDNA library “High Colon Metastasis Tissue”                 | 36984                               |
| 18              | Normal Colon – UC#3 Patient PCR (OligodT) cDNA library “Normal Colon Tumor Tissue”  | 36216                               |
| 19              | Colon Tumor – UC#3 Patient PCR (OligodT) cDNA library “High Colon Tumor Tissue”   | 41388                               |
| 20              | Liver Metastasis from Colon Tumor of UC#3 Patient PCR (OligodT) cDNA library “High Colon Metastasis Tissue”                 | 30956                               |
| 21              | G RRpz Human Prostate Cell Line   | 164801                              |
| 22              | WOca Human Prostate Cancer Cell Line  | 162088                              |

The KM12L4 and KM12C cell lines are described in Example 1 above. The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) 53:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as

models for the study of human breast and lung cancer (see, e.g., Chandrasekaran et al., *Cancer Res.* (1979) 39:870 (MDA-MB-231 and MCF-7); Gastpar et al., *J Med Chem* (1998) 41:4965 (MDA-MB-231 and MCF-7); Ranson et al., *Br J Cancer* (1998) 77:1586 (MDA-MB-231 and MCF-7); Kuang et al., *Nucleic Acids Res* (1998) 26:1116 (MDA-MB-231 and MCF-7); Varki et al., *Int J Cancer* (1987) 40:46 (UCP-3); Varki et al., *Tumour Biol.* (1990) 11:327; (MV-522 and UCP-3); Varki et al., *Anticancer Res.* (1990) 10:637; (MV-522); Kelner et al., *Anticancer Res* (1995) 15:867 (MV-522); and Zhang et al., *Anticancer Drugs* (1997) 8:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-treated HMEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMEC were prepared by incubation with 20ng/ml VEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation. The GRRpz cell line refers to low passage (3 passages or fewer) human prostate cells, and the WOca cell line refers to low passage (3 passages or fewer) human prostate cancer cells.

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters. The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac et al., *Genomics* (1996) 37(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought together computationally. These groups of clones are termed "clusters". Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the "purity" of each cluster can be controlled. For example, artifacts of clustering may occur in computational clustering just as artifacts can occur in "wet-lab" screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same

cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1<sup>st</sup>), and the determining the number of cDNA clones corresponding to the selected cluster in the second library (Clones in 2<sup>nd</sup>). Differential expression of the selected cluster in the first library relative to the second library is expressed as a “ratio” of percent expression between the two libraries. In general, the “ratio” is calculated by: 1) calculating the percent expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the “number of clones” corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the “depth” of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5, where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar, Biostatistical Analysis, Prentice Hall, Inc., USA, “Differences between Proportions,” pp 296-298 (1974)).

25

#### EXAMPLE 5

##### POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL BREAST CANCER CELLS VERSUS LOW METASTATIC BREAST CANCER CELLS

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential breast cancer tissue and low metastatic breast cancer cells. Expression of these sequences in breast cancer can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process.

A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential breast cancer cells and low metastatic potential breast cancer cells.

Table 7

Differentially expressed polynucleotides: Higher expression in high metastatic potential breast cancer (lib3) relative to low metastatic breast cancer cells (lib4)

| SEQ ID NOs: | Lib3 clones | Lib4 clones | lib3/lib4 |
|-------------|-------------|-------------|-----------|
| 472         | 64          | 0           | 62        |
| 1851        | 6           | 0           | 6         |
| 1856        | 8           | 0           | 8         |
| 1867        | 6           | 0           | 6         |
| 1872        | 6           | 0           | 6         |
| 1875        | 12          | 3           | 4         |
| 1923        | 89          | 22          | 4         |
| 2118        | 7           | 0           | 7         |
| 2119        | 7           | 0           | 7         |
| 2135        | 37          | 13          | 3         |
| 2190        | 19          | 0           | 19        |
| 2193        | 16          | 5           | 3         |
| 2232        | 12          | 2           | 6         |
| 2239        | 6           | 0           | 6         |
| 2338        | 21          | 2           | 10        |
| 2378        | 16          | 4           | 4         |
| 2394        | 6           | 0           | 6         |
| 2395        | 6           | 0           | 6         |
| 2490        | 13          | 3           | 4         |
| 2505        | 16          | 2           | 8         |
| 2540        | 8           | 1           | 8         |



| SEQ ID NOs: | Lib3 clones | Lib4 clones | lib3/lib4 |
|-------------|-------------|-------------|-----------|
| 2542        | 11          | 1           | 11        |
| 2607        | 11          | 2           | 5         |
| 2640        | 22          | 5           | 4         |
| 2674        | 8           | 0           | 8         |
| 2679        | 19          | 0           | 19        |
| 2684        | 14          | 4           | 3         |
| 2707        | 8           | 0           | 8         |
| 2724        | 9           | 0           | 9         |
| 2757        | 6           | 0           | 6         |
| 2776        | 10          | 0           | 10        |
| 2804        | 13          | 2           | 6         |
| 2818        | 6           | 0           | 6         |
| 2906        | 14          | 0           | 14        |
| 2959        | 26          | 8           | 3         |
| 2964        | 17          | 4           | 4         |
| 2968        | 6           | 0           | 6         |
| 2977        | 22          | 3           | 7         |
| 2980        | 13          | 1           | 13        |
| 3010        | 6           | 0           | 6         |
| 3043        | 10          | 1           | 10        |
| 3071        | 33          | 12          | 3         |
| 3072        | 9           | 1           | 9         |
| 3095        | 19          | 3           | 6         |
| 3097        | 11          | 2           | 5         |
| 3173        | 12          | 2           | 6         |
| 3203        | 8           | 1           | 8         |
| 3210        | 27          | 8           | 3         |
| 3212        | 13          | 1           | 13        |
| 3284        | 8           | 0           | 8         |
| 3288        | 6           | 0           | 6         |
| 3331        | 14          | 3           | 5         |
| 3335        | 13          | 1           | 13        |

Table 8

Differentially expressed polynucleotides: Higher expression in low metastatic breast cancer cells (lib4) relative to high metastatic potential breast cancer (lib3)

| SEQ ID NOs: | Lib 3 Clones | Lib 4 Clones | lib4/lib3 |
|-------------|--------------|--------------|-----------|
| 402         | 0            | 6            | 6         |
| 614         | 3            | 21           | 7         |
| 624         | 0            | 6            | 6         |
| 626         | 0            | 8            | 8         |
| 712         | 0            | 9            | 9         |
| 744         | 0            | 7            | 7         |
| 1325        | 2            | 29           | 15        |
| 1452        | 2            | 13           | 7         |
| 1880        | 0            | 9            | 9         |
| 1915        | 0            | 7            | 7         |
| 1951        | 0            | 6            | 6         |
| 1955        | 8            | 32           | 4         |
| 2015        | 0            | 7            | 7         |
| 2046        | 0            | 7            | 7         |
| 2076        | 1            | 22           | 23        |
| 2087        | 0            | 6            | 6         |
| 2124        | 0            | 9            | 9         |
| 2145        | 0            | 8            | 8         |
| 2162        | 0            | 6            | 6         |
| 2163        | 0            | 12           | 12        |
| 2164        | 5            | 19           | 4         |
| 2172        | 2            | 15           | 8         |
| 2192        | 5            | 16           | 3         |
| 2244        | 20           | 43           | 2         |
| 2266        | 3            | 18           | 6         |
| 2313        | 24           | 56           | 2         |
| 2346        | 1            | 13           | 13        |
| 2355        | 0            | 10           | 10        |
| 2371        | 0            | 6            | 6         |
| 2393        | 1            | 17           | 17        |
| 2404        | 1            | 21           | 22        |
| 2443        | 0            | 6            | 6         |
| 2460        | 0            | 11           | 11        |
| 2523        | 0            | 6            | 6         |
| 2575        | 1            | 10           | 10        |
| 2578        | 0            | 6            | 6         |
| 2584        | 1            | 17           | 17        |
| 2590        | 0            | 6            | 6         |

| SEQ ID NOs: | Lib 3 Clones | Lib 4 Clones | lib4/lib3 |
|-------------|--------------|--------------|-----------|
| 2609        | 1            | 9            | 9         |
| 2632        | 5            | 24           | 5         |
| 2714        | 5            | 24           | 5         |
| 2728        | 0            | 6            | 6         |
| 2752        | 1            | 14           | 14        |
| 2794        | 4            | 15           | 4         |
| 2826        | 0            | 7            | 7         |
| 2987        | 5            | 15           | 3         |
| 3005        | 1            | 14           | 14        |
| 3009        | 20           | 58           | 3         |
| 3047        | 4            | 17           | 4         |
| 3057        | 2            | 17           | 9         |
| 3075        | 2            | 11           | 6         |
| 3076        | 0            | 6            | 6         |
| 3102        | 0            | 6            | 6         |
| 3128        | 15           | 52           | 4         |
| 3132        | 15           | 52           | 4         |
| 3142        | 0            | 6            | 6         |
| 3187        | 22           | 49           | 2         |
| 3253        | 23           | 96           | 4         |
| 3282        | 19           | 46           | 2         |
| 3285        | 20           | 40           | 2         |
| 3346        | 0            | 9            | 9         |

## EXAMPLE 6

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL LUNG  
CANCER CELLS VERSUS LOW METASTATIC LUNG CANCER CELLS

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential lung cancer cells and low metastatic lung cancer cells. Expression of these sequences in lung cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory

sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential lung cancer cells and low metastatic potential lung cancer cells:

10

Table 9

Differentially expressed polynucleotides: Higher expression in high metastatic potential lung cancer cells (lib8) relative to low metastatic lung cancer cells (lib9)

| SEQ ID NO: | Lib8 clones | Lib9 clones | lib8/lib9 |
|------------|-------------|-------------|-----------|
| 14         | 10          | 0           | 10        |
| 137        | 5           | 0           | 5         |
| 151        | 5           | 0           | 7         |
| 152        | 9           | 0           | 13        |
| 171        | 6           | 0           | 8         |
| 200        | 10          | 0           | 14        |
| 254        | 5           | 0           | 7         |
| 262        | 5           | 0           | 7         |
| 271        | 5           | 0           | 7         |
| 348        | 6           | 1           | 8         |
| 412        | 5           | 0           | 7         |
| 507        | 5           | 0           | 7         |
| 520        | 6           | 0           | 8         |
| 530        | 5           | 0           | 7         |
| 588        | 5           | 0           | 7         |
| 623        | 7           | 0           | 10        |
| 637        | 7           | 0           | 10        |
| 660        | 5           | 0           | 7         |
| 678        | 8           | 0           | 11        |
| 680        | 5           | 0           | 7         |
| 700        | 9           | 2           | 6         |
| 714        | 28          | 13          | 3         |
| 774        | 11          | 0           | 15        |
| 812        | 5           | 0           | 7         |

| SEQ ID NO: | Lib8 clones | Lib9 clones | lib8/lib9 |
|------------|-------------|-------------|-----------|
| 834        | 8           | 2           | 6         |
| 901        | 11          | 2           | 8         |
| 1168       | 5           | 0           | 7         |
| 1333       | 6           | 0           | 8         |
| 1352       | 5           | 0           | 7         |
| 1524       | 11          | 1           | 15        |
| 1706       | 5           | 0           | 7         |
| 1752       | 17          | 9           | 3         |
| 1768       | 20          | 4           | 7         |
| 1769       | 5           | 0           | 7         |
| 1780       | 6           | 0           | 8         |
| 1781       | 40          | 3           | 19        |
| 1799       | 6           | 1           | 8         |
| 1803       | 6           | 1           | 8         |
| 1811       | 16          | 9           | 2         |
| 1884       | 6           | 0           | 8         |
| 1919       | 8           | 1           | 11        |
| 1939       | 6           | 0           | 8         |
| 1975       | 43          | 9           | 7         |
| 2024       | 12          | 1           | 17        |
| 2045       | 8           | 1           | 11        |
| 2060       | 20          | 13          | 2         |
| 2071       | 16          | 4           | 6         |
| 2128       | 5           | 0           | 7         |
| 2177       | 10          | 2           | 7         |
| 2181       | 44          | 13          | 5         |
| 2184       | 11          | 1           | 15        |
| 2185       | 10          | 4           | 3         |
| 2283       | 7           | 0           | 10        |
| 2311       | 10          | 4           | 3         |
| 2314       | 10          | 0           | 14        |
| 2393       | 14          | 6           | 3         |
| 2398       | 6           | 1           | 8         |
| 2460       | 10          | 4           | 3         |
| 2514       | 6           | 0           | 8         |
| 2597       | 5           | 0           | 7         |
| 2657       | 8           | 2           | 6         |
| 2669       | 6           | 1           | 8         |
| 2670       | 6           | 1           | 8         |
| 3047       | 21          | 3           | 10        |
| 3050       | 16          | 5           | 4         |
| 3092       | 7           | 1           | 10        |
| 3140       | 181         | 119         | 2         |
| 3157       | 5           | 0           | 7         |

| SEQ ID NO: | Lib8 clones | Lib9 clones | lib8/lib9 |
|------------|-------------|-------------|-----------|
| 3187       | 16          | 5           | 4         |
| 3210       | 5           | 0           | 7         |
| 3220       | 28          | 4           | 10        |
| 3236       | 7           | 1           | 10        |
| 3249       | 16          | 0           | 22        |
| 3264       | 8           | 2           | 6         |
| 3305       | 7           | 0           | 10        |
| 3309       | 20          | 0           | 28        |
| 3318       | 24          | 4           | 8         |
| 3330       | 5           | 0           | 7         |
| 3331       | 5           | 0           | 7         |

Table 10

Differentially expressed polynucleotides: Higher expression in low metastatic lung cancer cells (lib 9) relative to high metastatic potential lung cancer cells (lib 8)

| SEQ ID NO: | Lib 8 clones | Lib 9 clones | lib 9/lib 8 |
|------------|--------------|--------------|-------------|
| 24         | 3            | 20           | 5           |
| 53         | 0            | 18           | 13          |
| 64         | 0            | 8            | 6           |
| 70         | 0            | 11           | 8           |
| 105        | 10           | 66           | 5           |
| 129        | 0            | 16           | 11          |
| 214        | 1            | 14           | 10          |
| 233        | 4            | 35           | 6           |
| 237        | 0            | 13           | 9           |
| 264        | 0            | 29           | 21          |
| 329        | 2            | 17           | 6           |
| 368        | 1            | 37           | 26          |
| 370        | 0            | 11           | 8           |
| 418        | 0            | 8            | 6           |
| 450        | 0            | 9            | 6           |
| 461        | 0            | 9            | 6           |
| 484        | 0            | 26           | 19          |
| 494        | 0            | 41           | 29          |
| 517        | 1            | 12           | 9           |
| 522        | 1            | 11           | 8           |
| 581        | 1            | 17           | 12          |
| 614        | 3            | 23           | 5           |
| 706        | 0            | 11           | 8           |
| 726        | 5            | 23           | 3           |

| SEQ ID NO: | Lib 8 clones | Lib 9 clones | lib 9/lib 8 |
|------------|--------------|--------------|-------------|
| 806        | 0            | 14           | 10          |
| 824        | 0            | 9            | 6           |
| 836        | 1            | 14           | 10          |
| 874        | 0            | 12           | 9           |
| 900        | 5            | 21           | 3           |
| 1017       | 2            | 14           | 5           |
| 1144       | 0            | 8            | 6           |
| 1154       | 0            | 12           | 9           |
| 1166       | 2            | 45           | 16          |
| 1170       | 1            | 13           | 9           |
| 1302       | 2            | 13           | 5           |
| 1326       | 1            | 13           | 9           |
| 1327       | 1            | 13           | 9           |
| 1367       | 0            | 12           | 9           |
| 1377       | 0            | 12           | 9           |
| 1437       | 2            | 18           | 6           |
| 1442       | 1            | 14           | 10          |
| 1466       | 0            | 13           | 9           |
| 1476       | 0            | 13           | 9           |
| 1495       | 0            | 8            | 6           |
| 1496       | 1            | 13           | 9           |
| 1664       | 38           | 253          | 5           |
| 1682       | 1            | 17           | 12          |
| 1687       | 0            | 9            | 6           |
| 1758       | 0            | 8            | 6           |
| 1817       | 4            | 18           | 3           |
| 1837       | 3            | 16           | 4           |
| 1845       | 3            | 23           | 5           |
| 1856       | 2            | 17           | 6           |
| 1910       | 1            | 18           | 13          |
| 2146       | 2            | 16           | 9           |
| 2156       | 0            | 9            | 6           |
| 2463       | 0            | 12           | 9           |
| 2724       | 10           | 38           | 3           |
| 2749       | 403          | 2000         | 4           |
| 2801       | 6            | 25           | 3           |
| 2993       | 3            | 18           | 4           |
| 3080       | 0            | 10           | 7           |
| 3107       | 3            | 23           | 5           |
| 3292       | 0            | 20           | 14          |
| 3324       | 110          | 548          | 4           |

## EXAMPLE 7

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL COLON  
CANCER CELLS VERSUS LOW METASTATIC COLON CANCER CELLS

5 A number of polynucleotide sequences have been identified that are differentially  
expressed between cells derived from high metastatic potential colon cancer cells and low  
metastatic colon cancer cells. Expression of these sequences in colon cancer tissue can  
provide diagnostic, prognostic and/or treatment information. For example, sequences that  
are highly expressed in the high metastatic potential cells can be indicative of increased  
10 expression of genes or regulatory sequences involved in the metastatic process. A patient  
sample displaying an increased level of one or more of these polynucleotides may thus  
warrant more aggressive treatment. In another example, sequences that display higher  
expression in the low metastatic potential cells can be associated with genes or regulatory  
sequences that inhibit metastasis, and thus the expression of these polynucleotides in a  
15 sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic  
marker, a prognostic marker, for risk assessment, patient treatment and the like. These  
polynucleotide sequences can also be used in combination with other known molecular  
and/or biochemical markers.

20 The following table summarizes identified polynucleotides with differential  
expression between high metastatic potential colon cancer cells and low metastatic  
potential colon cancer cells:

Table 11

25 Differentially expressed polynucleotides: Higher expression in low metastatic colon cancer  
cells (lib 2) relative to high metastatic potential colon cancer cells (lib 1)

| SEQ ID NOs: | Lib 1 clones | Lib 2 clones | lib 2/lib 1 |
|-------------|--------------|--------------|-------------|
| 429         | 0            | 9            | 10          |
| 1494        | 0            | 8            | 9           |
| 1923        | 34           | 114          | 4           |
| 1986        | 3            | 12           | 4           |
| 2018        | 0            | 9            | 10          |
| 2036        | 2            | 10           | 5           |
| 2049        | 8            | 25           | 3           |
| 2135        | 24           | 87           | 4           |
| 2146        | 2            | 16           | 9           |



| SEQ ID NOs: | Lib 1 clones | Lib 2 clones | lib 2/lib 1 |
|-------------|--------------|--------------|-------------|
| 2208        | 6            | 27           | 5           |
| 2215        | 2            | 11           | 6           |
| 2239        | 1            | 10           | 11          |
| 2307        | 2            | 12           | 6           |
| 2313        | 28           | 62           | 2           |
| 2357        | 5            | 14           | 3           |
| 2360        | 3            | 21           | 8           |
| 2362        | 0            | 6            | 6           |
| 2378        | 3            | 12           | 4           |
| 2569        | 3            | 20           | 7           |
| 2571        | 0            | 6            | 6           |
| 2588        | 54           | 172          | 3           |
| 2592        | 15           | 41           | 3           |
| 2611        | 0            | 6            | 6           |
| 2636        | 0            | 9            | 10          |
| 2641        | 7            | 20           | 3           |
| 2650        | 0            | 9            | 10          |
| 2662        | 0            | 9            | 10          |
| 2674        | 4            | 13           | 4           |
| 2682        | 0            | 6            | 6           |
| 2702        | 9            | 25           | 3           |
| 2704        | 8            | 23           | 3           |
| 2715        | 2            | 12           | 6           |
| 2804        | 9            | 22           | 3           |
| 2821        | 13           | 29           | 2           |
| 2840        | 1            | 8            | 9           |
| 2846        | 2            | 15           | 8           |
| 2866        | 0            | 6            | 6           |
| 2906        | 0            | 6            | 6           |
| 2915        | 44           | 109          | 3           |
| 2933        | 0            | 6            | 6           |
| 2935        | 5            | 16           | 3           |
| 2957        | 1            | 11           | 12          |
| 2959        | 3            | 27           | 10          |
| 2977        | 16           | 30           | 2           |
| 2980        | 12           | 27           | 2           |
| 3000        | 2            | 13           | 7           |
| 3009        | 12           | 29           | 3           |
| 3115        | 0            | 7            | 8           |
| 3156        | 502          | 2170         | 5           |
| 3210        | 2            | 21           | 11          |
| 3211        | 0            | 9            | 10          |
| 3213        | 0            | 7            | 8           |
| 3235        | 2            | 12           | 6           |

| SEQ ID NOs: | Lib 1 clones | Lib 2 clones | lib 2/lib 1 |
|-------------|--------------|--------------|-------------|
| 3251        | 2            | 12           | 6           |
| 3296        | 3            | 12           | 4           |
| 3335        | 1            | 8            | 9           |

## EXAMPLE 8

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL  
COLON CANCER PATIENT TISSUE VERSUS NORMAL PATIENT TISSUE

5

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the advanced disease state which involves processes such as angiogenesis, dedifferentiation, cell replication, and metastasis. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment.

15 The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

20 The following tables summarize polynucleotides that are differentially expressed between high metastatic potential colon cancer tissue and normal colon tissue:

Table 12

Differentially expressed polynucleotides isolated from samples from two patients (patient 2 and patient 3 and) : Lower expression in high metastatic potential colon tissue (patient 2:lib 17; patient 3:lib 20) vs. normal colon tissue (patient 2:lib 15; patient 3:lib 18)

| SEQ ID NO: | lib 15 clones | lib 17 clones | lib 15/lib 17 |
|------------|---------------|---------------|---------------|
| 69         | 19            | 7             | 3             |
| 123        | 6             | 0             | 6             |
| 140        | 24            | 8             | 3             |
| 197        | 6             | 0             | 6             |
| 198        | 113           | 0             | 121           |

| SEQ ID NO: | lib 15 clones | lib 17 clones | lib 15/lib 17 |
|------------|---------------|---------------|---------------|
| 254        | 28            | 9             | 3             |
| 412        | 28            | 9             | 3             |
| 512        | 11            | 1             | 12            |
| 641        | 17            | 7             | 3             |
| 642        | 7             | 0             | 8             |
| 954        | 12            | 3             | 4             |
| 1011       | 209           | 16            | 14            |
| 1024       | 8             | 0             | 9             |
| 1040       | 12            | 3             | 4             |
| 1055       | 26            | 7             | 4             |
| 1106       | 31            | 15            | 2             |
| 1125       | 17            | 0             | 18            |
| 1129       | 17            | 0             | 18            |
| 1138       | 109           | 0             | 117           |
| 1244       | 14            | 1             | 15            |
| 1253       | 73            | 0             | 78            |
| 1283       | 34            | 7             | 5             |
| 1285       | 34            | 7             | 5             |
| 1339       | 13            | 4             | 3             |
| 1474       | 73            | 0             | 78            |
| 1505       | 18            | 3             | 6             |
| 1553       | 68            | 6             | 12            |
| 1554       | 2542          | 14            | 195           |
| 1605       | 2542          | 14            | 195           |
| 1628       | 6             | 0             | 6             |
| 1643       | 142           | 4             | 38            |
| 1753       | 12            | 0             | 10            |
| 1764       | 13            | 0             | 14            |
| SEQ ID NO: | Lib18 Clones  | Lib20 Clones  | lib18/lib20   |
| 105        | 28            | 11            | 2             |
| 198        | 21            | 0             | 18            |
| 254        | 9             | 0             | 8             |
| 412        | 9             | 0             | 8             |
| 1011       | 11            | 1             | 9             |
| 1138       | 14            | 0             | 12            |
| 1253       | 23            | 0             | 20            |
| 1643       | 18            | 0             | 15            |
| 1764       | 12            | 0             | 10            |
| 3156       | 140           | 43            | 3             |

Table 13

Differentially expressed polynucleotides isolated from samples from two patients (patient 2 and patient 3): Lower expression in normal colon tissue (patient 2:lib 15; patient 3:lib 18)vs. high metastatic potential colon tissue (patient 2:lib 17; patient 3:lib 20).

| SEQ ID NO: | Lib 15 Clones | Lib 17 Clones | lib 17/lib 15 |
|------------|---------------|---------------|---------------|
| 321        | 3             | 23            | 7             |
| 363        | 1             | 9             | 8             |
| 836        | 21            | 99            | 4             |
| 859        | 6             | 20            | 3             |
| 885        | 13            | 28            | 2             |
| 916        | 13            | 28            | 2             |
| 981        | 2             | 11            | 5             |
| 1226       | 8             | 70            | 8             |
| 1308       | 0             | 8             | 7             |
| 1317       | 29            | 84            | 3             |
| 1429       | 27            | 127           | 4             |
| 1442       | 0             | 9             | 8             |
| 1534       | 1             | 12            | 11            |
| 1540       | 12            | 43            | 3             |
| 1552       | 0             | 7             | 7             |
| 1556       | 1             | 9             | 8             |
| 1557       | 1             | 9             | 8             |
| 1569       | 2189          | 5122          | 2             |
| 1571       | 6             | 18            | 3             |
| 1576       | 3             | 25            | 8             |
| 1581       | 4             | 22            | 5             |
| 1601       | 25            | 157           | 6             |
| 1613       | 9             | 48            | 5             |
| 1616       | 15            | 61            | 4             |
| 1620       | 2             | 17            | 8             |
| 1622       | 4             | 99            | 23            |
| 1626       | 6             | 35            | 5             |
| 1647       | 4             | 22            | 5             |
| 1664       | 4             | 28            | 7             |
| 1683       | 2             | 18            | 8             |
| 1704       | 3             | 15            | 5             |
| 1800       | 0             | 7             | 7             |
| 2749       | 23            | 60            | 2             |
| 2784       | 4             | 14            | 3             |
| 2805       | 1             | 9             | 8             |
| 2976       | 3             | 14            | 4             |
| 3128       | 18            | 57            | 3             |
| 3129       | 26            | 124           | 4             |

| SEQ ID NO: | Lib 15 Clones | Lib 17 Clones | lib 17/lib 15 |
|------------|---------------|---------------|---------------|
| 3146       | 64            | 210           | 3             |
| 3150       | 940           | 2267          | 2             |
| 3151       | 2             | 15            | 7             |
|            |               |               |               |
| SEQ ID NO: | lib 18 clones | lib 20 clones | lib 20/lib 18 |
| 865        | 0             | 5             | 6             |
| 1569       | 1             | 7             | 8             |
| 1580       | 1             | 7             | 8             |
| 1590       | 1             | 7             | 8             |
| 2790       | 0             | 5             | 6             |

### EXAMPLE 9

#### POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH COLON TUMOR POTENTIAL PATIENT TISSUE VERSUS METASTASIZED COLON CANCER PATIENT TISSUE

5 A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from colon cancer tissue and cells derived from colon cancer tissue metastases to liver. Expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information associated with the transformation of precancerous tissue to malignant tissue. This information can be useful  
10 in the prevention of achieving the advanced malignant state in these tissues, and can be important in risk assessment for a patient.

The following table summarizes identified polynucleotides with differential expression between high tumor potential colon cancer tissue and cells derived from high metastatic potential colon cancer cells:

15

Table 14

Differentially expressed polynucleotides:  
Greater expression in metastatic colon tumor tissue (lib 20) vs.  
colon tumor tissue (lib 19)

| SEQ ID NO: | lib 19 clones | lib 20 clones | lib 20/lib 19 |
|------------|---------------|---------------|---------------|
| 937        | 0             | 6             | 8             |
| 976        | 0             | 5             | 7             |
| 1520       | 1             | 8             | 11            |
| 1546       | 1             | 11            | 15            |
| 1550       | 1             | 11            | 15            |
| 1574       | 1             | 8             | 11            |
| 1580       | 0             | 7             | 9             |
| 1590       | 0             | 7             | 9             |
| 1599       | 8             | 21            | 4             |
| 1607       | 158           | 632           | 5             |
| 1622       | 1             | 7             | 9             |

5

Table 15

Greater expression in colon tumor tissue (lib 19) than metastatic colon tissue (lib 20)

| SEQ ID NO: | lib 19 clones | lib 20 clones | lib 19/lib 20 |
|------------|---------------|---------------|---------------|
| 105        | 64            | 11            | 4             |
| 1011       | 53            | 1             | 40            |
| 1226       | 18            | 4             | 3             |
| 1571       | 8             | 0             | 6             |
| 1726       | 15            | 3             | 4             |
| 1811       | 17            | 2             | 6             |
| 2749       | 47            | 6             | 6             |
| 3146       | 19            | 2             | 7             |
| 3324       | 20            | 1             | 15            |

## EXAMPLE 10

10 POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH TUMOR POTENTIAL  
COLON CANCER PATIENT TISSUE VERSUS NORMAL PATIENT TISSUE

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high tumor potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can provide diagnostic,

prognostic and/or treatment information associated with the prevention of the malignant state in these tissues, and can be important in risk assessment for a patient. For example, sequences that are highly expressed in the potential colon cancer cells are associated with or can be indicative of increased expression of genes or regulatory sequences involved in early tumor progression. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential colon cancer cells and normal colon cells:

10

Table 16

Differentially expressed polynucleotides detected in samples from patient (patient 2)  
Higher expression in normal colon tissue (patient 2, lib 15)  
vs. tumor potential colon tissue (patient 2:lib16)

| SEQ ID NO: | lib 15 clones | lib 16 clones | lib 16/lib 15 |
|------------|---------------|---------------|---------------|
| 69         | 19            | 7             | 3             |
| 105        | 116           | 54            | 2             |
| 140        | 24            | 4             | 6             |
| 197        | 6             | 0             | 6             |
| 198        | 113           | 3             | 40            |
| 254        | 28            | 6             | 5             |
| 412        | 28            | 6             | 5             |
| 642        | 7             | 0             | 7             |
| 830        | 10            | 2             | 5             |
| 938        | 31            | 13            | 3             |
| 1011       | 209           | 37            | 6             |
| 1095       | 12            | 3             | 4             |
| 1125       | 17            | 0             | 18            |
| 1129       | 17            | 0             | 18            |
| 1138       | 109           | 1             | 115           |
| 1253       | 73            | 1             | 77            |
| 1283       | 34            | 13            | 3             |
| 1285       | 34            | 13            | 3             |
| 1339       | 13            | 3             | 5             |
| 1453       | 11            | 3             | 4             |
| 1474       | 73            | 1             | 77            |
| 1505       | 18            | 6             | 3             |
| 1554       | 2542          | 448           | 6             |
| 1605       | 2542          | 448           | 6             |

| SEQ ID NO: | lib 15 clones | lib 16 clones | lib 16/lib 15 |
|------------|---------------|---------------|---------------|
| 1614       | 36            | 14            | 3             |
| 1630       | 24            | 9             | 3             |
| 1643       | 142           | 2             | 75            |
| 1646       | 39            | 14            | 3             |
| 1649       | 24            | 8             | 3             |
| 1677       | 19            | 6             | 3             |
| 1753       | 13            | 0             | 14            |
| 1764       | 13            | 0             | 14            |
| 1766       | 177           | 65            | 3             |
| 1772       | 24            | 8             | 3             |

Table 17

Differentially expressed polypeptides detected in samples from patient. Lower expression in normal colon tissue (lib 18) than colon tumor tissue (lib 19)

| SEQ ID NO: | lib 18 clones | lib 19 clones | lib 19/lib 18 |
|------------|---------------|---------------|---------------|
| 3146       | 3             | 19            | 6             |
| 3150       | 21            | 228           | 10            |
| 3324       | 3             | 20            | 6             |



Table 18

Differentially expressed polypeptides detected in samples from patient. Higher expression in normal colon tissue (lib 18) than colon tumor tissue (lib 19)

| SEQ ID NO: | lib 18 clones | lib 19 clones | lib 18/lib 19 |
|------------|---------------|---------------|---------------|
| 198        | 21            | 2             | 12            |
| 465        | 6             | 0             | 7             |
| 489        | 6             | 0             | 7             |
| 745        | 6             | 0             | 7             |
| 859        | 11            | 2             | 6             |
| 976        | 7             | 0             | 8             |
| 1011       | 209           | 37            | 6             |
| 1045       | 8             | 1             | 9             |
| 1138       | 14            | 0             | 16            |
| 1253       | 23            | 0             | 26            |
| 1392       | 16            | 4             | 5             |
| 1474       | 23            | 0             | 26            |
| 1589       | 6             | 0             | 7             |
| 1591       | 22            | 11            | 2             |
| 1607       | 386           | 158           | 3             |
| 1643       | 18            | 0             | 21            |
| 1753       | 12            | 0             | 14            |
| 1764       | 12            | 0             | 14            |
|            |               |               |               |
| SEQ ID NO: | lib 18 clones | lib 19 clones | lib 19/lib 18 |
| 105        | 28            | 64            | 2             |
| 1011       | 11            | 53            | 4             |
| 1226       | 2             | 18            | 8             |
| 1251       | 6             | 19            | 3             |
| 1559       | 1             | 9             | 8             |
| 1571       | 0             | 8             | 7             |
| 1608       | 1             | 9             | 8             |
| 1766       | 2             | 13            | 6             |
| 1782       | 1             | 9             | 8             |
| 1811       | 1             | 17            | 15            |

Table 19

Differentially expressed polynucleotides:

Higher expression in colon tumor tissue

(patient 2, lib 16) vs. normal colon tissue (patient 2, lib 15)

| SEQ ID NO: | lib 15 clones | lib 16 clones | lib 16/lib 15 |
|------------|---------------|---------------|---------------|
| 7          | 1             | 9             | 9             |
| 164        | 6             | 19            | 3             |
| 734        | 4             | 15            | 4             |
| 836        | 21            | 53            | 2             |
| 928        | 2             | 11            | 5             |
| 965        | 2             | 11            | 5             |
| 987        | 2             | 11            | 5             |
| 1026       | 7             | 19            | 3             |
| 1044       | 4             | 16            | 4             |
| 1119       | 4             | 16            | 4             |
| 1226       | 8             | 46            | 5             |
| 1227       | 0             | 9             | 9             |
| 1251       | 7             | 95            | 13            |
| 1316       | 0             | 6             | 6             |
| 1429       | 27            | 81            | 3             |
| 1442       | 0             | 9             | 9             |
| 1540       | 12            | 28            | 2             |
| 1553       | 68            | 590           | 8             |
| 1560       | 4             | 24            | 6             |
| 1577       | 1             | 10            | 9             |
| 1588       | 5             | 20            | 4             |
| 1610       | 3             | 13            | 4             |
| 1620       | 2             | 23            | 11            |
| 1626       | 6             | 23            | 4             |
| 1673       | 2             | 15            | 7             |
| 2416       | 0             | 7             | 7             |
| 2749       | 23            | 54            | 2             |
| 2976       | 3             | 14            | 4             |
| 3129       | 26            | 64            | 2             |
| 3132       | 18            | 54            | 3             |

## EXAMPLE 11

## POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN GROWTH FACTOR-STIMULATED HUMAN MICROVASCULAR ENDOTHELIAL CELLS (HMEC) RELATIVE TO UNTREATED HMEC

A number of polynucleotide sequences have been identified that are differentially expressed between human microvascular endothelial cells (HMEC) that have been treated with growth factors relative to untreated HMEC.

Sequences that are differentially expressed between growth factor-treated HMEC and untreated HMEC can represent sequences encoding gene products involved in angiogenesis, metastasis (cell migration), and other developmental and oncogenic processes. For example, sequences that are more highly expressed in HMEC treated with growth factors (such as bFGF or VEGF) relative to untreated HMEC can serve as markers of cancer cells of higher metastatic potential. Detection of expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The following table summarizes identified polynucleotides with differential expression between growth factor-treated and untreated HMEC.

Table 20

Differentially expressed polynucleotides:

Higher expression in untreated HMEC (lib 12) vs. bFGF treated HMEC (lib 13)

| SEQ ID NO: | lib 12 clones | lib 13 clones | lib 12/lib 13 |
|------------|---------------|---------------|---------------|
| 849        | 6             | 0             | 6             |
| 1059       | 6             | 0             | 6             |
| 1206       | 12            | 2             | 6             |
| 3208       | 12            | 0             | 12            |

Lower expression in untreated HMEC (lib 12) vs. bFGF treated HMEC (lib 13)

|      |   |    |   |
|------|---|----|---|
| 2748 | 3 | 12 | 4 |
| 3325 | 0 | 6  | 6 |

Table 21

Differentially expressed polynucleotides:

Higher expression in untreated HMEC (lib 12) VEGF treated HMEC (lib14)

| SEQ ID NO: | lib 12 clones | lib 14 clones | lib 12/lib 14 |
|------------|---------------|---------------|---------------|
| 1150       | 9             | 0             | 9             |

5

Lower expression in untreated HMEC (lib 12) vs. VEGF treated HMEC (lib14)

|      |    |    |   |
|------|----|----|---|
| 3324 | 22 | 50 | 2 |
|------|----|----|---|

## EXAMPLE 12

### 10 POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN NORMAL PROSTATE CELLS RELATIVE TO PROSTATE CANCER CELLS

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from normal prostate cells and prostate cancer cells. Expression of these sequences prostate tissue suspected of being cancerous can provide diagnostic, prognostic and/or treatment information. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers. The following table summarizes identified polynucleotides with differential expression between high metastatic potential colon cancer cells and low metastatic potential colon cancer cells:

20

Table 22

Differentially expressed polynucleotides: normal prostate cell line (lib 21)  
vs. prostate cancer cell line (lib 22)

Higher in lib 21

| SEQ ID NO: | lib 21 clones | lib 22 clones | lib 21/lib 22 |
|------------|---------------|---------------|---------------|
| 53         | 17            | 2             | 8             |
| 1754       | 22            | 8             | 3             |

|      |    |   |    |
|------|----|---|----|
| 1801 | 7  | 0 | 7  |
| 1845 | 22 | 6 | 4  |
| 446  | 8  | 0 | 8  |
| 1410 | 6  | 0 | 6  |
| 2060 | 18 | 6 | 3  |
| 2143 | 12 | 3 | 4  |
| 2632 | 13 | 1 | 13 |
| 2899 | 16 | 2 | 8  |
| 3338 | 12 | 2 | 6  |

## Higher in lib 22

|      |     |     |    |
|------|-----|-----|----|
| 86   | 2   | 13  | 7  |
| 93   | 0   | 9   | 9  |
| 687  | 0   | 9   | 9  |
| 1269 | 1   | 15  | 15 |
| 1581 | 25  | 74  | 3  |
| 1647 | 25  | 74  | 3  |
| 1649 | 12  | 27  | 2  |
| 1710 | 5   | 16  | 3  |
| 1717 | 5   | 16  | 3  |
| 1772 | 12  | 27  | 2  |
| 1960 | 0   | 6   | 6  |
| 2987 | 0   | 6   | 6  |
| 3128 | 13  | 42  | 3  |
| 3132 | 13  | 42  | 3  |
| 3150 | 263 | 962 | 4  |
| 3222 | 0   | 6   | 6  |
| 3268 | 0   | 6   | 6  |

5

## EXAMPLE 13

## POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED ACROSS MULTIPLE LIBRARIES

A number of polynucleotide sequences have been identified that are differentially expressed between cancerous cells and normal cells across two or more tissue types tested (*i.e.*, breast, colon, lung, and prostate). Expression of these sequences in a tissue of any origin can provide diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. These polynucleotides can also serve as non-tissue specific markers of, for example, risk of metastasis of a tumor. The following polynucleotides were

differentially expressed but without tissue type-specificity in at least two of the breast, colon, lung, and prostate libraries tested: 53, 105, 355, 412, 614, 836, 1442, 1581, 1647, 1649, 1664, 1772, 1782, 1811, 1845, 1856, 1875, 1923, 2060, 2071, 2135, 2146, 2239, 2313, 2378, 2393, 2416, 2460, 2490, 2632, 2674, 2704, 2724, 2749, 2784, 2804, 2959, 2976, 2977, 2980, 2987, 3009, 3047, 3128, 3129, 3132, 3146, 3150, 3156, 3210, 3324, 3331, and 3335.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

#### Deposit Information:

The following materials were deposited with the American Type Culture Collection (ATCC); CMCC = Chiron Master Culture Collection:

#### 25 cDNA Libraries Deposited with ATCC

| Tube Number | Deposit Date | ATCC<br>Accession No. | CMCC<br>Accession No. |
|-------------|--------------|-----------------------|-----------------------|
| ES137       | May 30, 2000 |                       |                       |
| ES138       | May 30, 2000 |                       |                       |
| ES139       | May 30, 2000 |                       |                       |
| ES140       | May 30, 2000 |                       |                       |
| ES141       | May 30, 2000 |                       |                       |
| ES142       | May 30, 2000 |                       |                       |
| ES143       | May 30, 2000 |                       |                       |

| Tube Number | Deposit Date | ATCC<br>Accession No. | CMCC<br>Accession No. |
|-------------|--------------|-----------------------|-----------------------|
| ES137       | May 30, 2000 |                       |                       |
| ES144       | May 30, 2000 |                       |                       |
| ES145       | May 30, 2000 |                       |                       |
| ES146       | May 30, 2000 |                       |                       |
| ES147       | May 30, 2000 |                       |                       |
| ES148       | May 30, 2000 |                       |                       |
| ES149       | May 30, 2000 |                       |                       |
| ES150       | May 30, 2000 |                       |                       |
| ES151       | May 30, 2000 |                       |                       |
| ES152       | May 30, 2000 |                       |                       |
| ES153       | May 30, 2000 |                       |                       |
| ES154       | May 30, 2000 |                       |                       |
| ES155       | May 30, 2000 |                       |                       |
| ES156       | May 30, 2000 |                       |                       |
| ES157       | May 30, 2000 |                       |                       |
| ES158       | May 30, 2000 |                       |                       |
| ES159       | May 30, 2000 |                       |                       |
| ES160       | May 30, 2000 |                       |                       |
| ES161       | May 30, 2000 |                       |                       |
| ES162       | May 30, 2000 |                       |                       |
| ES163       | May 30, 2000 |                       |                       |
| ES164       | May 30, 2000 |                       |                       |
| ES165       | May 30, 2000 |                       |                       |
| ES166       | May 30, 2000 |                       |                       |
| ES167       | May 30, 2000 |                       |                       |

Table 23 lists the clones for each deposit, designated as “tube” number. This deposit is provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby.

#### 10 Retrieval of Individual Clones from Deposit of Pooled Clones

Where the ATCC deposit is composed of a pool of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones were then deposited as a pool of equal mixtures in the composite deposit. Particular clones

can be obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a  $T_m$  of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C). Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

Table 23

| Clone Name     | Tube   |
|----------------|--------|
| M00001351A:B02 | ES 137 |
| M00001356A:H11 | ES 137 |
| M00001363D:D09 | ES 137 |
| M00001395D:H02 | ES 137 |
| M00001439C:H06 | ES 137 |
| M00001476B:G10 | ES 137 |
| M00001582A:E02 | ES 137 |
| M00003750D:E06 | ES 137 |
| M00003761C:F02 | ES 137 |
| M00003770A:E05 | ES 137 |
| M00003786A:A11 | ES 137 |
| M00003800A:F09 | ES 137 |
| M00003816D:E11 | ES 137 |
| M00003902A:C03 | ES 137 |
| M00003991C:F06 | ES 137 |
| M00003995B:E03 | ES 137 |
| M00004046C:A08 | ES 137 |
| M00004105D:D05 | ES 137 |
| M00004139B:B10 | ES 137 |
| M00004140D:C03 | ES 137 |
| M00004144A:H05 | ES 137 |
| M00004152A:C12 | ES 137 |
| M00004155D:A10 | ES 137 |
| M00004168A:G11 | ES 137 |
| M00004197B:H10 | ES 137 |
| M00004222C:E03 | ES 137 |

| Clone Name     | Tube   |
|----------------|--------|
| M00004234A:E07 | ES 137 |
| M00004239B:F11 | ES 137 |
| M00004241B:H07 | ES 137 |
| M00004264B:A05 | ES 137 |
| M00004278A:F09 | ES 137 |
| M00004282D:C11 | ES 137 |
| M00004308C:C06 | ES 137 |
| M00004340C:C07 | ES 137 |
| M00004354D:E05 | ES 137 |
| M00004361A:H02 | ES 137 |
| M00004372B:F07 | ES 137 |
| M00004378A:B10 | ES 137 |
| M00004393B:E07 | ES 137 |
| M00023282A:C02 | ES 137 |
| M00023300D:C11 | ES 137 |
| M00023316C:G08 | ES 137 |
| M00023333D:C12 | ES 137 |
| M00023352B:F03 | ES 137 |
| M00023352D:H03 | ES 137 |
| M00023376B:G04 | ES 137 |
| M00023377B:F01 | ES 137 |
| M00023398B:D12 | ES 137 |
| M00023399C:E10 | ES 137 |
| M00026803A:F08 | ES 137 |
| M00026843B:D10 | ES 137 |
| M00026850D:F09 | ES 137 |



| Clone Name     | Tube   |
|----------------|--------|
| M00026851B:F01 | ES 137 |
| M00026856D:F02 | ES 137 |
| M00026857D:G12 | ES 137 |
| M00026859D:D01 | ES 137 |
| M00026860B:C05 | ES 137 |
| M00026865B:A06 | ES 137 |
| M00026868C:E11 | ES 137 |
| M00026878A:F05 | ES 137 |
| M00026882D:G09 | ES 137 |
| M00026885A:H09 | ES 137 |
| M00026901A:G07 | ES 137 |
| M00026914A:H10 | ES 137 |
| M00026915B:C06 | ES 137 |
| M00026918B:D01 | ES 137 |
| M00026922C:B02 | ES 137 |
| M00026922C:G03 | ES 137 |
| M00026926A:E10 | ES 137 |
| M00026927D:F02 | ES 137 |
| M00026928D:A03 | ES 137 |
| M00026935C:B04 | ES 137 |
| M00026941D:A04 | ES 137 |
| M00026944B:E03 | ES 137 |
| M00026946A:F12 | ES 137 |
| M00026980A:D09 | ES 137 |
| M00027016A:B06 | ES 137 |
| M00027018A:C09 | ES 137 |
| M00027021A:G02 | ES 137 |
| M00027022D:G11 | ES 137 |
| M00027030C:H06 | ES 137 |
| M00027035D:C06 | ES 137 |
| M00027049B:F05 | ES 137 |
| M00027078A:B02 | ES 137 |
| M00027080A:B01 | ES 137 |
| M00027085C:E11 | ES 137 |
| M00027094A:B03 | ES 137 |
| M00027103B:A09 | ES 137 |
| M00027108C:B03 | ES 137 |
| M00027121D:C05 | ES 137 |
| M00027135A:B11 | ES 137 |
| M00027136C:C09 | ES 137 |
| M00027141C:H03 | ES 137 |
| M00027159D:F03 | ES 137 |
| M00027162B:F05 | ES 137 |
| M00027178B:G09 | ES 137 |
| M00027179D:E06 | ES 138 |
| M00027181D:A05 | ES 138 |
| M00027195C:E04 | ES 138 |

| Clone Name     | Tube   |
|----------------|--------|
| M00027198B:B08 | ES 138 |
| M00027200A:F02 | ES 138 |
| M00027207B:F07 | ES 138 |
| M00027212D:E03 | ES 138 |
| M00027228D:A01 | ES 138 |
| M00027232D:B08 | ES 138 |
| M00027233B:C01 | ES 138 |
| M00027236A:E04 | ES 138 |
| M00027237C:B08 | ES 138 |
| M00027248A:C02 | ES 138 |
| M00027256B:H09 | ES 138 |
| M00027258A:A07 | ES 138 |
| M00027263A:F10 | ES 138 |
| M00027292D:F10 | ES 138 |
| M00027297A:C04 | ES 138 |
| M00027299B:B12 | ES 138 |
| M00027301A:G05 | ES 138 |
| M00027301B:B08 | ES 138 |
| M00027314C:D09 | ES 138 |
| M00027319D:B11 | ES 138 |
| M00027324D:C05 | ES 138 |
| M00027347C:G07 | ES 138 |
| M00027355A:B07 | ES 138 |
| M00027359B:G05 | ES 138 |
| M00027366A:F11 | ES 138 |
| M00027379C:B07 | ES 138 |
| M00027392B:H02 | ES 138 |
| M00027396D:G08 | ES 138 |
| M00027398C:F07 | ES 138 |
| M00027438C:G07 | ES 138 |
| M00027462A:D07 | ES 138 |
| M00027462B:H07 | ES 138 |
| M00027468A:C09 | ES 138 |
| M00027475B:E10 | ES 138 |
| M00027476A:C09 | ES 138 |
| M00027486A:F06 | ES 138 |
| M00027520A:C05 | ES 138 |
| M00027525B:D06 | ES 138 |
| M00027526D:F03 | ES 138 |
| M00027528C:B10 | ES 138 |
| M00027537C:B01 | ES 138 |
| M00027546C:B10 | ES 138 |
| M00027591B:C04 | ES 138 |
| M00027596A:A10 | ES 138 |
| M00027596C:E06 | ES 138 |
| M00027602B:C01 | ES 138 |
| M00027615A:F10 | ES 138 |

| Clone Name     | Tube   |
|----------------|--------|
| M00027617B:C12 | ES 138 |
| M00027620D:F11 | ES 138 |
| M00027625A:H01 | ES 138 |
| M00027634A:D11 | ES 138 |
| M00027641C:A03 | ES 138 |
| M00027647C:D03 | ES 138 |
| M00027652B:F11 | ES 138 |
| M00027668C:H12 | ES 138 |
| M00027729D:H06 | ES 138 |
| M00027733A:A02 | ES 138 |
| M00027741B:F09 | ES 138 |
| M00027743A:C03 | ES 138 |
| M00027801C:C11 | ES 138 |
| M00027813C:F01 | ES 138 |
| M00027818C:C07 | ES 138 |
| M00027836D:F12 | ES 138 |
| M00027837C:D09 | ES 138 |
| M00028120D:F12 | ES 138 |
| M00028066C:D07 | ES 138 |
| M00028184D:G10 | ES 138 |
| M00028185B:A06 | ES 138 |
| M00028196D:A03 | ES 138 |
| M00028201B:H12 | ES 138 |
| M00028207D:E09 | ES 138 |
| M00028210B:D02 | ES 138 |
| M00028212C:B08 | ES 138 |
| M00028215D:F03 | ES 138 |
| M00028220A:B04 | ES 138 |
| M00028314D:F05 | ES 138 |
| M00028316B:H12 | ES 138 |
| M00028354A:B12 | ES 138 |
| M00028354D:A03 | ES 138 |
| M00028357A:G10 | ES 138 |
| M00028362A:G11 | ES 138 |
| M00028364C:G08 | ES 138 |
| M00028369D:E08 | ES 138 |
| M00028617C:A12 | ES 138 |
| M00028768C:D05 | ES 138 |
| M00028770A:D04 | ES 138 |
| M00028772C:B09 | ES 138 |
| M00028775D:F03 | ES 138 |
| M00028777B:G12 | ES 138 |
| M00031368A:E10 | ES 138 |
| M00031417C:G09 | ES 138 |
| M00031419D:C04 | ES 138 |
| M00031485D:G02 | ES 138 |
| M00032480B:E10 | ES 139 |

| Clone Name     | Tube   |
|----------------|--------|
| M00032492A:C01 | ES 139 |
| M00032495B:D02 | ES 139 |
| M00032499C:A01 | ES 139 |
| M00032508B:H03 | ES 139 |
| M00032510D:F12 | ES 139 |
| M00032510D:G06 | ES 139 |
| M00032513D:F01 | ES 139 |
| M00032530D:C02 | ES 139 |
| M00032535D:H01 | ES 139 |
| M00032539B:C11 | ES 139 |
| M00032540A:A09 | ES 139 |
| M00032541D:H08 | ES 139 |
| M00032545B:H09 | ES 139 |
| M00032545D:G05 | ES 139 |
| M00032550D:C02 | ES 139 |
| M00032551B:G05 | ES 139 |
| M00032577A:C04 | ES 139 |
| M00032578A:G06 | ES 139 |
| M00032584A:H08 | ES 139 |
| M00032592A:H11 | ES 139 |
| M00032597C:B01 | ES 139 |
| M00032638C:G08 | ES 139 |
| M00032638D:A06 | ES 139 |
| M00032668D:G12 | ES 139 |
| M00032678C:D06 | ES 139 |
| M00032688D:D11 | ES 139 |
| M00032712B:G02 | ES 139 |
| M00032724A:C05 | ES 139 |
| M00032725C:F06 | ES 139 |
| M00032726C:C01 | ES 139 |
| M00032731B:C10 | ES 139 |
| M00032731C:C07 | ES 139 |
| M00032737B:E09 | ES 139 |
| M00032739A:A06 | ES 139 |
| M00032744B:F10 | ES 139 |
| M00032766B:D12 | ES 139 |
| M00032766C:A04 | ES 139 |
| M00032790B:A07 | ES 139 |
| M00032793A:F06 | ES 139 |
| M00032797B:G02 | ES 139 |
| M00032808B:G10 | ES 139 |
| M00032811B:D02 | ES 139 |
| M00032829B:E06 | ES 139 |
| M00032830D:G03 | ES 139 |
| M00032831C:G07 | ES 139 |
| M00032853D:G12 | ES 139 |
| M00032864B:B09 | ES 139 |

| Clone Name     | Tube   |
|----------------|--------|
| M00032871D:E11 | ES 139 |
| M00032876C:D06 | ES 139 |
| M00032907A:G04 | ES 139 |
| M00032909A:B06 | ES 139 |
| M00032917D:G09 | ES 139 |
| M00032918B:D08 | ES 139 |
| M00032918B:E06 | ES 139 |
| M00032918C:B10 | ES 139 |
| M00032921B:H08 | ES 139 |
| M00032933A:C10 | ES 139 |
| M00032939B:E07 | ES 139 |
| M00032940A:C02 | ES 139 |
| M00032942D:C12 | ES 139 |
| M00032944B:B02 | ES 139 |
| M00032984C:G05 | ES 139 |
| M00032990B:A11 | ES 139 |
| M00032994A:A08 | ES 139 |
| M00032995C:C05 | ES 139 |
| M00033007C:E01 | ES 139 |
| M00033019B:E10 | ES 139 |
| M00033033C:H01 | ES 139 |
| M00033034C:A06 | ES 139 |
| M00033034C:F02 | ES 139 |
| M00033037D:C11 | ES 139 |
| M00033074A:C08 | ES 139 |
| M00033130B:F06 | ES 139 |
| M00033140D:F06 | ES 139 |
| M00033173D:C01 | ES 139 |
| M00033176B:E12 | ES 139 |
| M00033186C:D11 | ES 139 |
| M00033189D:F08 | ES 139 |
| M00033202D:G06 | ES 139 |
| M00033204B:A07 | ES 139 |
| M00033205A:F03 | ES 139 |
| M00033217B:H07 | ES 139 |
| M00033218A:C04 | ES 139 |
| M00033223B:H07 | ES 139 |
| M00033226A:A11 | ES 139 |
| M00033231D:B09 | ES 139 |
| M00033231D:G10 | ES 139 |
| M00033243B:A05 | ES 139 |
| M00033246C:E08 | ES 139 |
| M00033248A:B02 | ES 139 |
| M00033261C:D12 | ES 139 |
| M00033262D:A11 | ES 139 |
| M00033263B:G04 | ES 139 |
| M00033276B:G08 | ES 139 |

| Clone Name     | Tube   |
|----------------|--------|
| M00033185C:D01 | ES 139 |
| M00033288B:D12 | ES 140 |
| M00033300D:H12 | ES 140 |
| M00033306D:G08 | ES 140 |
| M00033306D:H09 | ES 140 |
| M00033308B:G05 | ES 140 |
| M00033343C:H08 | ES 140 |
| M00033345D:A09 | ES 140 |
| M00033346C:A05 | ES 140 |
| M00033347C:F02 | ES 140 |
| M00033349D:F05 | ES 140 |
| M00033358A:H12 | ES 140 |
| M00033362C:C05 | ES 140 |
| M00033375A:G04 | ES 140 |
| M00033376A:C12 | ES 140 |
| M00033377D:A05 | ES 140 |
| M00033410B:C09 | ES 140 |
| M00033424B:A04 | ES 140 |
| M00033424D:H12 | ES 140 |
| M00033425A:C10 | ES 140 |
| M00033427D:F01 | ES 140 |
| M00033432B:H10 | ES 140 |
| M00033437C:A07 | ES 140 |
| M00033437C:C03 | ES 140 |
| M00033442A:D06 | ES 140 |
| M00033446C:G08 | ES 140 |
| M00033446D:B02 | ES 140 |
| M00033450C:A02 | ES 140 |
| M00033451A:H01 | ES 140 |
| M00033454A:D09 | ES 140 |
| M00033457D:A05 | ES 140 |
| M00033560D:G07 | ES 140 |
| M00033561C:A02 | ES 140 |
| M00033566C:E08 | ES 140 |
| M00033570B:C08 | ES 140 |
| M00033570B:E06 | ES 140 |
| M00033570C:C10 | ES 140 |
| M00033578D:G02 | ES 140 |
| M00033581C:H10 | ES 140 |
| M00033581D:D08 | ES 140 |
| M00033583B:E06 | ES 140 |
| M00033583D:B05 | ES 140 |
| M00033584D:G11 | ES 140 |
| M00033585D:A02 | ES 140 |
| M00033588C:G04 | ES 140 |
| M00033594C:B03 | ES 140 |
| M00033595A:C11 | ES 140 |

| Clone Name     | Tube   |
|----------------|--------|
| M00038259A:G08 | ES 140 |
| M00038259B:A02 | ES 140 |
| M00038259B:G08 | ES 140 |
| M00038259C:H09 | ES 140 |
| M00038272A:G01 | ES 140 |
| M00038272D:F11 | ES 140 |
| M00038279C:A11 | ES 140 |
| M00038284B:H04 | ES 140 |
| M00038303A:C03 | ES 140 |
| M00038303C:D02 | ES 140 |
| M00038303D:E07 | ES 140 |
| M00038315C:G11 | ES 140 |
| M00038325D:F12 | ES 140 |
| M00038326B:D04 | ES 140 |
| M00038327A:C11 | ES 140 |
| M00038327D:A05 | ES 140 |
| M00038328D:A03 | ES 140 |
| M00038329A:E08 | ES 140 |
| M00038387B:A07 | ES 140 |
| M00038614C:H11 | ES 140 |
| M00038615A:H12 | ES 140 |
| M00038616D:B12 | ES 140 |
| M00038618C:C08 | ES 140 |
| M00038619B:A03 | ES 140 |
| M00038620B:E09 | ES 140 |
| M00038631C:B10 | ES 140 |
| M00038631D:B02 | ES 140 |
| M00038632C:B09 | ES 140 |
| M00038633A:D07 | ES 140 |
| M00038633B:G02 | ES 140 |
| M00038635A:G09 | ES 140 |
| M00038635B:C08 | ES 140 |
| M00038638D:H03 | ES 140 |
| M00038639B:C03 | ES 140 |
| M00038639D:F07 | ES 140 |
| M00038661A:A07 | ES 140 |
| M00038662B:A12 | ES 140 |
| M00038663B:H06 | ES 140 |
| M00038663D:H10 | ES 140 |
| M00038664C:E04 | ES 140 |
| M00038991A:D01 | ES 140 |
| M00038994A:A10 | ES 140 |
| M00038995C:G08 | ES 140 |
| M00038995D:E05 | ES 140 |
| M00038999B:G11 | ES 140 |
| M00038999D:C11 | ES 140 |
| M00039002D:G11 | ES 140 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039004B:A06 | ES 140 |
| M00039004B:C11 | ES 140 |
| M00039005C:H01 | ES 141 |
| M00039006D:B01 | ES 141 |
| M00039011D:C10 | ES 141 |
| M00039013A:C09 | ES 141 |
| M00039013D:F02 | ES 141 |
| M00039014A:H10 | ES 141 |
| M00039014B:C04 | ES 141 |
| M00039015A:D07 | ES 141 |
| M00039015B:G10 | ES 141 |
| M00039015B:H09 | ES 141 |
| M00039015D:H04 | ES 141 |
| M00039016A:A02 | ES 141 |
| M00039016D:G06 | ES 141 |
| M00039024B:B10 | ES 141 |
| M00039025A:H09 | ES 141 |
| M00039026D:F05 | ES 141 |
| M00039028C:B11 | ES 141 |
| M00039030B:E02 | ES 141 |
| M00039036C:B05 | ES 141 |
| M00039039B:E03 | ES 141 |
| M00039039B:F09 | ES 141 |
| M00039042B:B02 | ES 141 |
| M00039043B:E01 | ES 141 |
| M00039049D:G07 | ES 141 |
| M00039050A:H10 | ES 141 |
| M00039052C:F07 | ES 141 |
| M00039058A:A04 | ES 141 |
| M00039058C:H02 | ES 141 |
| M00039059C:G08 | ES 141 |
| M00039061B:F08 | ES 141 |
| M00039063B:D08 | ES 141 |
| M00039064D:H09 | ES 141 |
| M00039066D:G08 | ES 141 |
| M00039068B:B04 | ES 141 |
| M00039068C:E06 | ES 141 |
| M00039070D:C02 | ES 141 |
| M00039072C:C03 | ES 141 |
| M00039072C:E02 | ES 141 |
| M00039079A:A05 | ES 141 |
| M00039080C:H06 | ES 141 |
| M00039081B:G06 | ES 141 |
| M00039082B:A05 | ES 141 |
| M00039084C:G07 | ES 141 |
| M00039084C:H03 | ES 141 |
| M00039084C:H04 | ES 141 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039084D:D07 | ES 141 |
| M00039096A:A05 | ES 141 |
| M00039096A:E07 | ES 141 |
| M00039097D:D06 | ES 141 |
| M00039099A:H08 | ES 141 |
| M00039104D:C09 | ES 141 |
| M00039105C:B08 | ES 141 |
| M00039107C:E04 | ES 141 |
| M00039108D:B06 | ES 141 |
| M00039112B:C05 | ES 141 |
| M00039118B:C05 | ES 141 |
| M00039118D:A06 | ES 141 |
| M00039120C:C09 | ES 141 |
| M00039120C:H03 | ES 141 |
| M00039123A:B10 | ES 141 |
| M00039124C:F03 | ES 141 |
| M00039124C:H02 | ES 141 |
| M00039124C:H08 | ES 141 |
| M00039126D:A08 | ES 141 |
| M00039127A:G11 | ES 141 |
| M00039127D:E10 | ES 141 |
| M00039129C:D04 | ES 141 |
| M00039133B:F08 | ES 141 |
| M00039135D:F05 | ES 141 |
| M00039135D:G02 | ES 141 |
| M00039135D:H02 | ES 141 |
| M00039139A:C09 | ES 141 |
| M00039139C:G12 | ES 141 |
| M00039140A:B08 | ES 141 |
| M00039140D:A04 | ES 141 |
| M00039140D:D09 | ES 141 |
| M00039141C:E01 | ES 141 |
| M00039142D:B11 | ES 141 |
| M00039144C:E06 | ES 141 |
| M00039147A:F10 | ES 141 |
| M00039156A:B11 | ES 141 |
| M00039158B:G12 | ES 141 |
| M00039166B:G06 | ES 141 |
| M00039167B:H09 | ES 141 |
| M00039168C:A04 | ES 141 |
| M00039169A:E12 | ES 141 |
| M00039170A:B10 | ES 141 |
| M00039170C:F05 | ES 141 |
| M00039171B:D11 | ES 141 |
| M00039177B:D03 | ES 141 |
| M00039179A:G09 | ES 141 |
| M00039180A:A07 | ES 141 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039196B:H06 | ES 141 |
| M00039196D:A07 | ES 141 |
| M00039200A:C10 | ES 141 |
| M00039211A:C12 | ES 141 |
| M00039212C:C12 | ES 142 |
| M00039213A:D01 | ES 142 |
| M00039213B:F05 | ES 142 |
| M00039218A:F03 | ES 142 |
| M00039221A:H03 | ES 142 |
| M00039224A:E12 | ES 142 |
| M00039228A:B05 | ES 142 |
| M00039230A:A10 | ES 142 |
| M00039230D:D09 | ES 142 |
| M00039230D:G12 | ES 142 |
| M00039233A:A03 | ES 142 |
| M00039238A:B12 | ES 142 |
| M00039238D:A08 | ES 142 |
| M00039241A:E11 | ES 142 |
| M00039249A:C12 | ES 142 |
| M00039249C:G11 | ES 142 |
| M00039255C:E12 | ES 142 |
| M00039257D:C03 | ES 142 |
| M00039258B:E06 | ES 142 |
| M00039258D:B08 | ES 142 |
| M00039260C:G03 | ES 142 |
| M00039263D:A12 | ES 142 |
| M00039266A:B02 | ES 142 |
| M00039266D:F12 | ES 142 |
| M00039266D:H04 | ES 142 |
| M00039273B:F02 | ES 142 |
| M00039273D:B02 | ES 142 |
| M00039274B:G07 | ES 142 |
| M00039276B:H09 | ES 142 |
| M00039277D:G10 | ES 142 |
| M00039279B:C11 | ES 142 |
| M00039279B:H02 | ES 142 |
| M00039279C:B08 | ES 142 |
| M00039281D:B04 | ES 142 |
| M00039284D:B12 | ES 142 |
| M00039286A:C06 | ES 142 |
| M00039287C:A06 | ES 142 |
| M00039288C:B11 | ES 142 |
| M00039293A:H04 | ES 142 |
| M00039293B:C11 | ES 142 |
| M00039295B:D03 | ES 142 |
| M00039297C:H08 | ES 142 |
| M00039298B:B06 | ES 142 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039298B:D03 | ES 142 |
| M00039298D:B04 | ES 142 |
| M00039299B:G12 | ES 142 |
| M00039300C:C09 | ES 142 |
| M00039300C:G04 | ES 142 |
| M00039301B:F06 | ES 142 |
| M00039303C:F11 | ES 142 |
| M00039304D:B09 | ES 142 |
| M00039308B:G08 | ES 142 |
| M00039310A:C07 | ES 142 |
| M00039313D:G04 | ES 142 |
| M00039316A:C01 | ES 142 |
| M00039318B:B09 | ES 142 |
| M00039319B:H12 | ES 142 |
| M00039319C:A04 | ES 142 |
| M00039322A:F04 | ES 142 |
| M00039328D:D07 | ES 142 |
| M00039329A:C01 | ES 142 |
| M00039329C:B10 | ES 142 |
| M00039333D:D09 | ES 142 |
| M00039334B:E03 | ES 142 |
| M00039335A:E08 | ES 142 |
| M00039339A:H07 | ES 142 |
| M00039339C:F03 | ES 142 |
| M00039340A:D05 | ES 142 |
| M00039340B:E07 | ES 142 |
| M00039340B:G08 | ES 142 |
| M00039341C:H11 | ES 142 |
| M00039341D:D07 | ES 142 |
| M00039343B:F12 | ES 142 |
| M00039344B:G07 | ES 142 |
| M00039345A:D09 | ES 142 |
| M00039345C:C12 | ES 142 |
| M00039381C:H08 | ES 142 |
| M00039381D:C02 | ES 142 |
| M00039384C:E02 | ES 142 |
| M00039384C:F08 | ES 142 |
| M00039385B:E09 | ES 142 |
| M00039391D:F08 | ES 142 |
| M00039396D:B04 | ES 142 |
| M00039397B:H09 | ES 142 |
| M00039398A:B10 | ES 142 |
| M00039401B:D02 | ES 142 |
| M00039402B:E03 | ES 142 |
| M00039403A:G12 | ES 142 |
| M00039404B:A05 | ES 142 |
| M00039407B:G02 | ES 142 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039411C:E07 | ES 142 |
| M00039412D:G06 | ES 142 |
| M00039414D:G03 | ES 142 |
| M00039415D:E01 | ES 142 |
| M00039417A:D03 | ES 142 |
| M00039417A:E12 | ES 142 |
| M00039417B:F01 | ES 143 |
| M00039417C:A01 | ES 143 |
| M00039417C:G01 | ES 143 |
| M00039418B:D08 | ES 143 |
| M00039420D:D03 | ES 143 |
| M00039422D:F04 | ES 143 |
| M00039425C:G01 | ES 143 |
| M00039425D:E12 | ES 143 |
| M00039428C:E01 | ES 143 |
| M00039430B:F12 | ES 143 |
| M00039431B:F04 | ES 143 |
| M00039432C:A01 | ES 143 |
| M00039444C:H02 | ES 143 |
| M00039452C:G09 | ES 143 |
| M00039454B:A11 | ES 143 |
| M00039455D:H04 | ES 143 |
| M00039456A:C08 | ES 143 |
| M00039458B:H11 | ES 143 |
| M00039461A:F04 | ES 143 |
| M00039465A:A08 | ES 143 |
| M00039472C:B08 | ES 143 |
| M00039475C:E10 | ES 143 |
| M00039476B:A02 | ES 143 |
| M00039477A:B03 | ES 143 |
| M00039477D:A10 | ES 143 |
| M00039611D:D11 | ES 143 |
| M00039612B:B10 | ES 143 |
| M00039612B:G05 | ES 143 |
| M00039616A:B10 | ES 143 |
| M00039616B:C01 | ES 143 |
| M00039619B:D02 | ES 143 |
| M00039631A:C10 | ES 143 |
| M00039633D:D05 | ES 143 |
| M00039636C:D11 | ES 143 |
| M00039637C:A10 | ES 143 |
| M00039652B:D05 | ES 143 |
| M00039655B:H09 | ES 143 |
| M00039655C:C07 | ES 143 |
| M00039655C:E08 | ES 143 |
| M00039660C:C10 | ES 143 |
| M00039663C:G09 | ES 143 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039664D:G07 | ES 143 |
| M00039672D:D10 | ES 143 |
| M00039673A:F09 | ES 143 |
| M00039675D:B03 | ES 143 |
| M00039675D:H05 | ES 143 |
| M00039677A:B08 | ES 143 |
| M00039681B:H09 | ES 143 |
| M00039682A:C08 | ES 143 |
| M00039682C:H11 | ES 143 |
| M00039684D:B08 | ES 143 |
| M00039685A:A08 | ES 143 |
| M00039686C:C05 | ES 143 |
| M00039686C:E06 | ES 143 |
| M00039688C:G06 | ES 143 |
| M00039689C:E08 | ES 143 |
| M00039696A:E05 | ES 143 |
| M00039697B:F11 | ES 143 |
| M00039700B:D02 | ES 143 |
| M00039702A:B12 | ES 143 |
| M00039702A:B02 | ES 143 |
| M00039705D:F02 | ES 143 |
| M00039707A:D02 | ES 143 |
| M00039710C:G03 | ES 143 |
| M00039720D:D02 | ES 143 |
| M00039727C:B09 | ES 143 |
| M00039729A:A10 | ES 143 |
| M00039771C:E11 | ES 143 |
| M00039773D:A09 | ES 143 |
| M00039773D:F11 | ES 143 |
| M00039774C:A03 | ES 143 |
| M00039774C:C09 | ES 143 |
| M00039775A:A09 | ES 143 |
| M00039777C:E05 | ES 143 |
| M00039778B:G03 | ES 143 |
| M00039778C:A04 | ES 143 |
| M00039781D:D10 | ES 143 |
| M00039782A:H10 | ES 143 |
| M00039785D:G05 | ES 143 |
| M00039788A:E03 | ES 143 |
| M00039788B:A06 | ES 143 |
| M00039788C:A01 | ES 143 |
| M00039790B:D03 | ES 143 |
| M00039792A:B04 | ES 143 |
| M00039793D:C05 | ES 143 |
| M00039794A:E04 | ES 143 |
| M00039795B:H10 | ES 143 |
| M00039795D:E10 | ES 143 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039795D:G06 | ES 143 |
| M00039797C:G05 | ES 143 |
| M00039798B:B02 | ES 143 |
| M00039799A:D10 | ES 143 |
| M00039801A:H11 | ES 143 |
| M00039807A:D01 | ES 143 |
| M00039808D:H02 | ES 143 |
| M00039810A:H10 | ES 143 |
| M00039813B:B01 | ES 144 |
| M00039813B:D11 | ES 144 |
| M00039815C:F09 | ES 144 |
| M00039816B:D04 | ES 144 |
| M00039816C:D05 | ES 144 |
| M00039820A:F11 | ES 144 |
| M00039820A:H11 | ES 144 |
| M00039820B:B06 | ES 144 |
| M00039827B:F07 | ES 144 |
| M00039828B:C05 | ES 144 |
| M00039832A:B12 | ES 144 |
| M00039835A:F07 | ES 144 |
| M00039838A:F05 | ES 144 |
| M00039839B:B01 | ES 144 |
| M00039839C:E05 | ES 144 |
| M00039847A:F06 | ES 144 |
| M00039851B:G11 | ES 144 |
| M00039851C:D12 | ES 144 |
| M00039854B:F09 | ES 144 |
| M00039855C:F01 | ES 144 |
| M00039857B:G10 | ES 144 |
| M00039859A:F06 | ES 144 |
| M00039859C:G10 | ES 144 |
| M00039864A:A07 | ES 144 |
| M00039866B:A08 | ES 144 |
| M00039869B:F06 | ES 144 |
| M00039875D:A10 | ES 144 |
| M00039876D:H09 | ES 144 |
| M00039877C:C03 | ES 144 |
| M00039879C:F05 | ES 144 |
| M00039879D:B11 | ES 144 |
| M00039880A:H11 | ES 144 |
| M00039884A:H11 | ES 144 |
| M00039885C:D01 | ES 144 |
| M00039887C:E07 | ES 144 |
| M00039887D:C04 | ES 144 |
| M00039888B:D03 | ES 144 |
| M00039890A:H05 | ES 144 |
| M00039894C:H07 | ES 144 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039896C:H01 | ES 144 |
| M00039897D:C10 | ES 144 |
| M00039898A:A08 | ES 144 |
| M00039898D:C06 | ES 144 |
| M00039903A:H07 | ES 144 |
| M00039903C:D01 | ES 144 |
| M00039903C:F03 | ES 144 |
| M00039909C:G05 | ES 144 |
| M00039909D:C02 | ES 144 |
| M00039910C:G10 | ES 144 |
| M00039914D:G12 | ES 144 |
| M00039915D:C11 | ES 144 |
| M00039927A:F04 | ES 144 |
| M00039928B:G05 | ES 144 |
| M00039936C:C05 | ES 144 |
| M00039938C:A08 | ES 144 |
| M00039938C:E11 | ES 144 |
| M00039940A:D07 | ES 144 |
| M00039940D:G08 | ES 144 |
| M00039973D:C08 | ES 144 |
| M00039973D:D12 | ES 144 |
| M00039975C:C11 | ES 144 |
| M00039976D:A12 | ES 144 |
| M00039978A:G03 | ES 144 |
| M00039981A:E08 | ES 144 |
| M00039982C:H04 | ES 144 |
| M00039983D:A06 | ES 144 |
| M00039984A:C02 | ES 144 |
| M00039984B:G12 | ES 144 |
| M00039984D:G12 | ES 144 |
| M00039987A:F09 | ES 144 |
| M00039987C:E12 | ES 144 |
| M00039987C:G08 | ES 144 |
| M00039988A:E06 | ES 144 |
| M00039990C:D10 | ES 144 |
| M00040004D:B03 | ES 144 |
| M00040005B:C11 | ES 144 |
| M00040005D:B07 | ES 144 |
| M00040007D:A06 | ES 144 |
| M00040009D:B07 | ES 144 |
| M00040010A:F10 | ES 144 |
| M00040014B:D01 | ES 144 |
| M00040014D:D10 | ES 144 |
| M00040014D:F03 | ES 144 |
| M00040015C:F08 | ES 144 |
| M00040016C:H12 | ES 144 |
| M00040017A:C06 | ES 144 |

| Clone Name     | Tube   |
|----------------|--------|
| M00040017D:G03 | ES 144 |
| M00040019A:E01 | ES 144 |
| M00040021A:F09 | ES 144 |
| M00040022C:D06 | ES 144 |
| M00040026B:F06 | ES 144 |
| M00040029A:B03 | ES 144 |
| M00040029A:G04 | ES 144 |
| M00040031A:E06 | ES 144 |
| M00040032A:B03 | ES 144 |
| M00040032A:D09 | ES 144 |
| M00040037A:E11 | ES 145 |
| M00040038D:G04 | ES 145 |
| M00040039D:D06 | ES 145 |
| M00040040A:A06 | ES 145 |
| M00040041C:C09 | ES 145 |
| M00040042B:A10 | ES 145 |
| M00040047C:F05 | ES 145 |
| M00040052D:F12 | ES 145 |
| M00040055D:A06 | ES 145 |
| M00040055D:B01 | ES 145 |
| M00040060C:H10 | ES 145 |
| M00040062B:B05 | ES 145 |
| M00040070B:B07 | ES 145 |
| M00040071B:A10 | ES 145 |
| M00040072C:G09 | ES 145 |
| M00040076C:D06 | ES 145 |
| M00040077D:C11 | ES 145 |
| M00040080C:C06 | ES 145 |
| M00040081C:E01 | ES 145 |
| M00040085D:A10 | ES 145 |
| M00040085D:E04 | ES 145 |
| M00040087D:F08 | ES 145 |
| M00040088C:E10 | ES 145 |
| M00040089A:G08 | ES 145 |
| M00040089B:E04 | ES 145 |
| M00040089C:E06 | ES 145 |
| M00040090B:G09 | ES 145 |
| M00040092B:F05 | ES 145 |
| M00040093B:C02 | ES 145 |
| M00040093D:D03 | ES 145 |
| M00040097A:C12 | ES 145 |
| M00040098C:B01 | ES 145 |
| M00040098D:E04 | ES 145 |
| M00040098D:G12 | ES 145 |
| M00040100C:E05 | ES 145 |
| M00040100D:B06 | ES 145 |
| M00040103B:H10 | ES 145 |



| Clone Name     | Tube   |
|----------------|--------|
| M00040105C:F11 | ES 145 |
| M00040106B:B09 | ES 145 |
| M00040107B:H07 | ES 145 |
| M00040111C:D05 | ES 145 |
| M00040115B:A04 | ES 145 |
| M00040115B:H12 | ES 145 |
| M00040118D:G10 | ES 145 |
| M00040121B:C05 | ES 145 |
| M00040122D:A02 | ES 145 |
| M00040123A:A09 | ES 145 |
| M00040124D:H01 | ES 145 |
| M00040129D:E10 | ES 145 |
| M00040302C:A04 | ES 145 |
| M00040304B:F06 | ES 145 |
| M00040305A:D11 | ES 145 |
| M00040305C:H06 | ES 145 |
| M00040307B:F01 | ES 145 |
| M00040307C:F10 | ES 145 |
| M00040309A:E11 | ES 145 |
| M00004825D:D05 | ES 145 |
| M00004832D:H02 | ES 145 |
| M00004839C:H02 | ES 145 |
| M00005018A:B05 | ES 145 |
| M00005297D:H08 | ES 145 |
| M00005308A:D06 | ES 145 |
| M00005351C:G05 | ES 145 |
| M00005352C:A02 | ES 145 |
| M00005358B:B06 | ES 145 |
| M00005359A:D04 | ES 145 |
| M00005379A:E04 | ES 145 |
| M00005382B:F08 | ES 145 |
| M00005384A:C11 | ES 145 |
| M00005402B:F08 | ES 145 |
| M00005445D:B01 | ES 145 |
| M00005449B:B10 | ES 145 |
| M00005449B:D01 | ES 145 |
| M00005457C:A03 | ES 145 |
| M00005458A:F11 | ES 145 |
| M00005498A:H06 | ES 145 |
| M00005531D:F06 | ES 145 |
| M00005539D:G01 | ES 145 |
| M00005555A:A10 | ES 145 |
| M00005556B:D02 | ES 145 |
| M00005601D:D08 | ES 145 |
| M00005614B:B01 | ES 145 |
| M00005623A:G02 | ES 145 |
| M00005623D:G12 | ES 145 |

| Clone Name     | Tube   |
|----------------|--------|
| M00005625A:C02 | ES 145 |
| M00005673B:B12 | ES 145 |
| M00005778B:F09 | ES 145 |
| M00005805D:D12 | ES 145 |
| M00005820C:E04 | ES 145 |
| M00006581D:F08 | ES 145 |
| M00006599D:B02 | ES 145 |
| M00006657C:G05 | ES 145 |
| M00006680B:D02 | ES 145 |
| M00006712C:H09 | ES 145 |
| M00006809B:B09 | ES 145 |
| M00006861B:F09 | ES 145 |
| M00006866A:D07 | ES 146 |
| M00006886D:H02 | ES 146 |
| M00006893C:E07 | ES 146 |
| M00006897A:H02 | ES 146 |
| M00006928D:D07 | ES 146 |
| M00006935C:F06 | ES 146 |
| M00006968A:G08 | ES 146 |
| M00006977C:G04 | ES 146 |
| M00006977D:A03 | ES 146 |
| M00007012D:H08 | ES 146 |
| M00007013A:D09 | ES 146 |
| M00007026B:H09 | ES 146 |
| M00007108B:A02 | ES 146 |
| M00007112C:B10 | ES 146 |
| M00007116C:G02 | ES 146 |
| M00007124D:H10 | ES 146 |
| M00007136A:A03 | ES 146 |
| M00007149A:G02 | ES 146 |
| M00007157C:F11 | ES 146 |
| M00007165B:G11 | ES 146 |
| M00007194A:B09 | ES 146 |
| M00007929C:B08 | ES 146 |
| M00007941D:C09 | ES 146 |
| M00007943D:C09 | ES 146 |
| M00007972B:H12 | ES 146 |
| M00007976A:C10 | ES 146 |
| M00007992C:F06 | ES 146 |
| M00007994A:G02 | ES 146 |
| M00008006B:B03 | ES 146 |
| M00008026B:C11 | ES 146 |
| M00008045A:H02 | ES 146 |
| M00008053A:F10 | ES 146 |
| M00008063B:A06 | ES 146 |

| Clone Name     | Tube   |
|----------------|--------|
| M00021665B:F12 | ES 146 |
| M00021671D:F12 | ES 146 |
| M00021852D:A05 | ES 146 |
| M00021866D:A03 | ES 146 |
| M00021908D:G12 | ES 146 |
| M00021919C:A10 | ES 146 |
| M00021923C:D11 | ES 146 |
| M00021955A:H02 | ES 146 |
| M00021964C:E10 | ES 146 |
| M00021972D:C11 | ES 146 |
| M00022005C:C06 | ES 146 |
| M00022015B:B07 | ES 146 |
| M00022054A:H03 | ES 146 |
| M00022084D:B01 | ES 146 |
| M00022099B:D06 | ES 146 |
| M00022105C:C12 | ES 146 |
| M00022127C:H03 | ES 146 |
| M00022135C:B05 | ES 146 |
| M00022138A:E05 | ES 146 |
| M00022175D:D12 | ES 146 |
| M00022178B:D06 | ES 146 |
| M00022181C:D01 | ES 146 |
| M00022183B:C02 | ES 146 |
| M00022184C:C11 | ES 146 |
| M00022233C:A12 | ES 146 |
| M00022234C:D06 | ES 146 |
| M00022247A:E02 | ES 146 |
| M00022257A:B09 | ES 146 |
| M00022262D:G03 | ES 146 |
| M00022264B:G10 | ES 146 |
| M00022363C:G12 | ES 146 |
| M00022365D:A03 | ES 146 |
| M00022373A:B05 | ES 146 |
| M00022373C:B07 | ES 146 |
| M00022391B:E01 | ES 146 |
| M00022391D:F10 | ES 146 |
| M00022416A:A07 | ES 146 |
| M00022421B:C11 | ES 146 |
| M00022433A:E02 | ES 146 |
| M00022434D:D06 | ES 146 |
| M00022440B:E01 | ES 146 |
| M00022444D:G01 | ES 146 |
| M00022467C:B12 | ES 146 |
| M00022489C:G04 | ES 146 |

| Clone Name     | Tube   |
|----------------|--------|
| M00022492C:A02 | ES 146 |
| M00022495D:H08 | ES 146 |
| M00022496B:E12 | ES 146 |
| M00022499A:B02 | ES 146 |
| M00022533A:A08 | ES 146 |
| M00022579C:C11 | ES 146 |
| M00022597D:A06 | ES 146 |
| M00022602A:E09 | ES 146 |
| M00022615D:G05 | ES 146 |
| M00022634D:C08 | ES 146 |
| M00022640C:C12 | ES 146 |
| M00022641C:H05 | ES 146 |
| M00022646A:H10 | ES 146 |
| M00022662D:G11 | ES 146 |
| M00022667D:B02 | ES 146 |
| M00022668B:B12 | ES 146 |
| M00022670D:H11 | ES 146 |
| M00022671B:A08 | ES 146 |
| M00022684A:C02 | ES 146 |
| M00022731A:D02 | ES 147 |
| M00022739A:B03 | ES 147 |
| M00022747D:E03 | ES 147 |
| M00022767B:G11 | ES 147 |
| M00022785C:G06 | ES 147 |
| M00022793D:B01 | ES 147 |
| M00022795B:G06 | ES 147 |
| M00022797B:G08 | ES 147 |
| M00022817A:H02 | ES 147 |
| M00022821C:C09 | ES 147 |
| M00022823C:C01 | ES 147 |
| M00022830D:D01 | ES 147 |
| M00022834B:G11 | ES 147 |
| M00022854A:B03 | ES 147 |
| M00022856C:A07 | ES 147 |
| M00022860C:G04 | ES 147 |
| M00022885C:H05 | ES 147 |
| M00022895A:H08 | ES 147 |
| M00022910A:A06 | ES 147 |
| M00022925C:A08 | ES 147 |
| M00022928B:C01 | ES 147 |
| M00022930C:E02 | ES 147 |
| M00022938B:F07 | ES 147 |
| M00022968B:E02 | ES 147 |
| M00022976C:F04 | ES 147 |

| Clone Name     | Tube   |
|----------------|--------|
| M00022979A:D05 | ES 147 |
| M00022986D:H09 | ES 147 |
| M00022997A:F06 | ES 147 |
| M00023001C:C08 | ES 147 |
| M00023003C:D07 | ES 147 |
| M00023007A:H04 | ES 147 |
| M00023007C:E10 | ES 147 |
| M00023020C:G08 | ES 147 |
| M00023024D:F12 | ES 147 |
| M00023032A:B05 | ES 147 |
| M00023039D:B05 | ES 147 |
| M00023042D:D02 | ES 147 |
| M00023044B:D02 | ES 147 |
| M00023094A:B11 | ES 147 |
| M00023100A:E12 | ES 147 |
| M00039181D:E05 | ES 147 |
| M00039184A:D03 | ES 147 |
| M00039184B:B09 | ES 147 |
| M00039361B:E01 | ES 147 |
| M00039363A:C09 | ES 147 |
| M00039366C:B07 | ES 147 |
| M00039367B:H02 | ES 147 |
| M00039371B:H06 | ES 147 |
| M00039372C:D12 | ES 147 |
| M00039374B:B07 | ES 147 |
| M00039374C:H12 | ES 147 |
| M00039374C:H02 | ES 147 |
| M00039376D:H07 | ES 147 |
| M00039377D:E12 | ES 147 |
| M00039378D:H07 | ES 147 |
| M00039379A:B03 | ES 147 |
| M00039380C:C09 | ES 147 |
| M00039482B:G02 | ES 147 |
| M00039493A:C04 | ES 147 |
| M00039496B:D08 | ES 147 |
| M00039496B:H09 | ES 147 |
| M00039497C:C06 | ES 147 |
| M00039499C:A04 | ES 147 |
| M00039500C:C04 | ES 147 |
| M00039505C:E03 | ES 147 |
| M00039508A:C12 | ES 147 |
| M00039508C:G01 | ES 147 |
| M00039510C:G02 | ES 147 |
| M00039512C:D06 | ES 147 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039515A:A06 | ES 147 |
| M00039515D:C11 | ES 147 |
| M00039517B:G12 | ES 147 |
| M00039521A:A02 | ES 147 |
| M00039521D:H03 | ES 147 |
| M00039528B:B12 | ES 147 |
| M00039529C:D07 | ES 147 |
| M00039530B:E02 | ES 147 |
| M00039533A:C12 | ES 147 |
| M00039533B:G08 | ES 147 |
| M00039533D:F04 | ES 147 |
| M00039535D:D10 | ES 147 |
| M00039536C:C10 | ES 147 |
| M00039536C:H11 | ES 147 |
| M00039561A:B07 | ES 147 |
| M00039561B:A09 | ES 147 |
| M00039562B:G02 | ES 147 |
| M00039564B:C01 | ES 147 |
| M00039570A:D10 | ES 147 |
| M00039570B:D10 | ES 147 |
| M00039584C:C01 | ES 147 |
| M00039584C:C11 | ES 147 |
| M00039587C:F12 | ES 147 |
| M00039590D:D02 | ES 147 |
| M00039591C:D06 | ES 147 |
| M00039595C:E05 | ES 147 |
| M00039597D:F04 | ES 147 |
| M00039600A:A11 | ES 148 |
| M00039604B:E05 | ES 148 |
| M00039604D:G03 | ES 148 |
| M00039606B:D08 | ES 148 |
| M00039607D:E08 | ES 148 |
| M00039608D:H01 | ES 148 |
| M00039609D:F07 | ES 148 |
| M00039624A:H09 | ES 148 |
| M00039624B:F12 | ES 148 |
| M00039625B:G08 | ES 148 |
| M00039626D:F04 | ES 148 |
| M00039629B:F01 | ES 148 |
| M00039629D:B04 | ES 148 |
| M00039630A:C08 | ES 148 |
| M00039630C:H04 | ES 148 |
| M00039641A:A05 | ES 148 |
| M00039641C:D07 | ES 148 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039642D:B12 | ES 148 |
| M00039642D:H09 | ES 148 |
| M00039643C:B04 | ES 148 |
| M00039645C:E01 | ES 148 |
| M00039647A:H11 | ES 148 |
| M00039736D:G08 | ES 148 |
| M00039740B:F10 | ES 148 |
| M00039752B:G08 | ES 148 |
| M00039755A:B08 | ES 148 |
| M00039756B:H06 | ES 148 |
| M00039760B:B08 | ES 148 |
| M00040131B:D11 | ES 148 |
| M00040131C:F03 | ES 148 |
| M00040131D:G08 | ES 148 |
| M00040133B:B03 | ES 148 |
| M00040136C:F08 | ES 148 |
| M00040138B:H03 | ES 148 |
| M00040141D:F05 | ES 148 |
| M00040143A:H05 | ES 148 |
| M00040145D:D03 | ES 148 |
| M00040147D:H11 | ES 148 |
| M00040160B:A10 | ES 148 |
| M00040162A:E01 | ES 148 |
| M00040169B:F08 | ES 148 |
| M00040173D:B05 | ES 148 |
| M00040174C:E10 | ES 148 |
| M00040174D:G03 | ES 148 |
| M00040181B:H09 | ES 148 |
| M00040181D:H10 | ES 148 |
| M00040182D:D06 | ES 148 |
| M00040183A:F07 | ES 148 |
| M00040184C:A11 | ES 148 |
| M00040191A:B09 | ES 148 |
| M00040221A:G11 | ES 148 |
| M00040222D:G02 | ES 148 |
| M00040223A:C05 | ES 148 |
| M00040226A:H10 | ES 148 |
| M00040230A:H02 | ES 148 |
| M00040231B:C08 | ES 148 |
| M00040232D:B07 | ES 148 |
| M00040233A:H02 | ES 148 |
| M00040233C:G05 | ES 148 |
| M00040252C:C06 | ES 148 |
| M00040253C:A05 | ES 148 |

| Clone Name     | Tube   |
|----------------|--------|
| M00040254B:C10 | ES 148 |
| M00040256A:A06 | ES 148 |
| M00040257D:H10 | ES 148 |
| M00040260B:D02 | ES 148 |
| M00040260C:D04 | ES 148 |
| M00040261C:F01 | ES 148 |
| M00040262B:B06 | ES 148 |
| M00040264D:G05 | ES 148 |
| M00040265D:B07 | ES 148 |
| M00040265D:C08 | ES 148 |
| M00040267A:E06 | ES 148 |
| M00040267C:C04 | ES 148 |
| M00040271B:E12 | ES 148 |
| M00040271C:D08 | ES 148 |
| M00040273B:H12 | ES 148 |
| M00040274A:D07 | ES 148 |
| M00040274A:H11 | ES 148 |
| M00040280C:H05 | ES 148 |
| M00040281D:B01 | ES 148 |
| M00040282A:A03 | ES 148 |
| M00040286C:C02 | ES 148 |
| M00040287C:B09 | ES 148 |
| M00040287D:D07 | ES 148 |
| M00039746C:A08 | ES 148 |
| M00039746C:G09 | ES 148 |
| M00039746C:H05 | ES 148 |
| M00039746C:H06 | ES 148 |
| M00039746D:D11 | ES 148 |
| M00039748A:F11 | ES 148 |
| M00039748C:F11 | ES 148 |
| M00039749D:D05 | ES 148 |
| M00039761D:E10 | ES 148 |
| M00039762B:F07 | ES 148 |
| M00039764C:D07 | ES 148 |
| M00039766A:G07 | ES 148 |
| M00039766D:H01 | ES 149 |
| M00039767B:A04 | ES 149 |
| M00039767C:E12 | ES 149 |
| M00039770A:G11 | ES 149 |
| M00039770C:E04 | ES 149 |
| M00039942D:C01 | ES 149 |
| M00039943B:F10 | ES 149 |
| M00039945C:F09 | ES 149 |
| M00039946B:F08 | ES 149 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039947A:D06 | ES 149 |
| M00039947C:G03 | ES 149 |
| M00039948A:E03 | ES 149 |
| M00039948D:D11 | ES 149 |
| M00039951A:B07 | ES 149 |
| M00039951B:B12 | ES 149 |
| M00039951B:C03 | ES 149 |
| M00039955C:C04 | ES 149 |
| M00039957C:C09 | ES 149 |
| M00039957D:A12 | ES 149 |
| M00039958A:A08 | ES 149 |
| M00039958C:B09 | ES 149 |
| M00040201C:G11 | ES 149 |
| M00040202A:F05 | ES 149 |
| M00040203A:H06 | ES 149 |
| M00040203B:A05 | ES 149 |
| M00040203D:H11 | ES 149 |
| M00040206A:A07 | ES 149 |
| M00040207B:D08 | ES 149 |
| M00040208A:C03 | ES 149 |
| M00040208B:A07 | ES 149 |
| M00040208D:G09 | ES 149 |
| M00040217D:B07 | ES 149 |
| M00040218C:C02 | ES 149 |
| M00040219B:D02 | ES 149 |
| M00040219D:E08 | ES 149 |
| M00040291D:C05 | ES 149 |
| M00040293D:G04 | ES 149 |
| M00040294D:D12 | ES 149 |
| M00040296D:E09 | ES 149 |
| M00040298B:G02 | ES 149 |
| M00040299B:F10 | ES 149 |
| M00040313C:D05 | ES 149 |
| M00040313D:E04 | ES 149 |
| M00040314D:H05 | ES 149 |
| M00040317A:H03 | ES 149 |
| M00040317D:F02 | ES 149 |
| M00040318A:B02 | ES 149 |
| M00040318C:H11 | ES 149 |
| M00040320D:F02 | ES 149 |
| M00040323B:C12 | ES 149 |
| M00040323C:G11 | ES 149 |
| M00040326A:F04 | ES 149 |
| M00040327B:G06 | ES 149 |

| Clone Name     | Tube   |
|----------------|--------|
| M00040332D:B05 | ES 149 |
| M00040333D:G05 | ES 149 |
| M00040334D:B02 | ES 149 |
| M00040334D:C07 | ES 149 |
| M00040342B:D12 | ES 149 |
| M00040345D:A09 | ES 149 |
| M00040346A:C11 | ES 149 |
| M00040347D:F09 | ES 149 |
| M00040349D:B09 | ES 149 |
| M00040351B:F02 | ES 149 |
| M00040351D:A11 | ES 149 |
| M00040364A:E05 | ES 149 |
| M00040366A:B01 | ES 149 |
| M00040368A:A12 | ES 149 |
| M00040368A:F01 | ES 149 |
| M00040368D:E09 | ES 149 |
| M00040371C:H05 | ES 149 |
| M00040375C:B06 | ES 149 |
| M00040376C:G02 | ES 149 |
| M00040377C:G07 | ES 149 |
| M00040383A:H02 | ES 149 |
| M00040383D:C04 | ES 149 |
| M00040385C:D02 | ES 149 |
| M00040386A:A02 | ES 149 |
| M00040387C:E07 | ES 149 |
| M00040387D:H05 | ES 149 |
| M00040390A:H02 | ES 149 |
| M00040390B:F02 | ES 149 |
| M00040391A:D10 | ES 149 |
| M00040392B:H01 | ES 149 |
| M00040392C:B12 | ES 149 |
| M00040394A:D04 | ES 149 |
| M00040395B:D11 | ES 149 |
| M00042534A:A05 | ES 149 |
| M00042538B:E06 | ES 149 |
| M00042543C:G04 | ES 149 |
| M00042558A:F03 | ES 149 |
| M00042560A:F12 | ES 149 |
| M00042565C:A08 | ES 149 |
| M00042566C:C05 | ES 149 |
| M00042567B:H10 | ES 149 |
| M00042693D:E04 | ES 149 |
| M00042696B:E05 | ES 149 |
| M00042697D:C07 | ES 150 |

| Clone Name     | Tube   |
|----------------|--------|
| M00042698D:D10 | ES 150 |
| M00042698D:E01 | ES 150 |
| M00042702B:G02 | ES 150 |
| M00042704A:F09 | ES 150 |
| M00042711B:A11 | ES 150 |
| M00042717A:C07 | ES 150 |
| M00042737C:H04 | ES 150 |
| M00042740A:E09 | ES 150 |
| M00042742D:D05 | ES 150 |
| M00042887C:D07 | ES 150 |
| M00042895A:D10 | ES 150 |
| M00042895C:G01 | ES 150 |
| M00042902D:B08 | ES 150 |
| M00042904B:E07 | ES 150 |
| M00042905A:F11 | ES 150 |
| M00042905B:C03 | ES 150 |
| M00042905D:D02 | ES 150 |
| M00042347D:H11 | ES 150 |
| M00042348B:E05 | ES 150 |
| M00042349D:D07 | ES 150 |
| M00042431B:G08 | ES 150 |
| M00042431C:F01 | ES 150 |
| M00042431D:C10 | ES 150 |
| M00042432D:E02 | ES 150 |
| M00042435A:A11 | ES 150 |
| M00042436B:H09 | ES 150 |
| M00042437A:D04 | ES 150 |
| M00042439B:B03 | ES 150 |
| M00042439B:D03 | ES 150 |
| M00042440B:E09 | ES 150 |
| M00042463A:F09 | ES 150 |
| M00042470C:E05 | ES 150 |
| M00042511A:H04 | ES 150 |
| M00042515C:F08 | ES 150 |
| M00042751C:C12 | ES 150 |
| M00042752A:E11 | ES 150 |
| M00042756B:F11 | ES 150 |
| M00042756D:A10 | ES 150 |
| M00042759B:G11 | ES 150 |
| M00042760A:C12 | ES 150 |
| M00042765C:D04 | ES 150 |
| M00042767B:G10 | ES 150 |
| M00042769C:E09 | ES 150 |
| M00042770B:B12 | ES 150 |

| Clone Name     | Tube   |
|----------------|--------|
| M00042770C:C04 | ES 150 |
| M00042771C:F06 | ES 150 |
| M00042774C:C05 | ES 150 |
| M00042781A:A07 | ES 150 |
| M00042784A:H06 | ES 150 |
| M00042788C:F11 | ES 150 |
| M00042790C:C07 | ES 150 |
| M00042792A:H01 | ES 150 |
| M00042797D:D10 | ES 150 |
| M00042799D:F08 | ES 150 |
| M00042800A:A03 | ES 150 |
| M00042802C:C04 | ES 150 |
| M00042806C:F07 | ES 150 |
| M00042807D:D05 | ES 150 |
| M00042823C:C02 | ES 150 |
| M00042830B:E02 | ES 150 |
| M00042839B:B11 | ES 150 |
| M00042841D:H07 | ES 150 |
| M00042849D:F11 | ES 150 |
| M00042852B:A03 | ES 150 |
| M00042852C:A01 | ES 150 |
| M00042856B:H02 | ES 150 |
| M00042352C:H03 | ES 150 |
| M00042352D:C01 | ES 150 |
| M00042352D:G09 | ES 150 |
| M00042448A:C09 | ES 150 |
| M00042448C:H12 | ES 150 |
| M00042453B:G09 | ES 150 |
| M00042518D:A06 | ES 150 |
| M00042518D:D04 | ES 150 |
| M00043296B:G09 | ES 150 |
| M00043304B:D05 | ES 150 |
| M00043304C:D02 | ES 150 |
| M00043305B:G02 | ES 150 |
| M00043306C:B03 | ES 150 |
| M00043306D:B07 | ES 150 |
| M00043310C:G06 | ES 150 |
| M00043311C:E03 | ES 150 |
| M00043312C:E08 | ES 150 |
| M00043320B:A07 | ES 150 |
| M00043324D:H11 | ES 150 |
| M00043328D:H02 | ES 150 |
| M00043332C:G04 | ES 150 |
| M00043334B:A10 | ES 150 |

| Clone Name     | Tube   |
|----------------|--------|
| M00043338B:A03 | ES 150 |
| M00043338B:C11 | ES 150 |
| M00043339A:F11 | ES 150 |
| M00043340B:H08 | ES 150 |
| M00043344D:E04 | ES 150 |
| M00043345C:A06 | ES 150 |
| M00043346A:G01 | ES 150 |
| M00043350D:B11 | ES 151 |
| M00043351D:A11 | ES 151 |
| M00043352D:B05 | ES 151 |
| M00043352D:C03 | ES 151 |
| M00043359B:D10 | ES 151 |
| M00043359C:G01 | ES 151 |
| M00043361B:A01 | ES 151 |
| M00043366A:A02 | ES 151 |
| M00043366C:H05 | ES 151 |
| M00043367B:A08 | ES 151 |
| M00043368C:F09 | ES 151 |
| M00043370B:C08 | ES 151 |
| M00043372C:G05 | ES 151 |
| M00043377A:C03 | ES 151 |
| M00043378A:H10 | ES 151 |
| M00043379D:H02 | ES 151 |
| M00043383C:F12 | ES 151 |
| M00043383D:A02 | ES 151 |
| M00043384B:B02 | ES 151 |
| M00043386A:B08 | ES 151 |
| M00043389C:E03 | ES 151 |
| M00043389D:D07 | ES 151 |
| M00043391A:C10 | ES 151 |
| M00043391A:G08 | ES 151 |
| M00043392D:C11 | ES 151 |
| M00043393A:B08 | ES 151 |
| M00043401D:G08 | ES 151 |
| M00043402C:D08 | ES 151 |
| M00043405A:D11 | ES 151 |
| M00043405C:G12 | ES 151 |
| M00043405C:G02 | ES 151 |
| M00043406B:G12 | ES 151 |
| M00043407C:E05 | ES 151 |
| M00043408B:D11 | ES 151 |
| M00043409B:B03 | ES 151 |
| M00043410C:A09 | ES 151 |
| M00043411B:D08 | ES 151 |

| Clone Name     | Tube   |
|----------------|--------|
| M00043411D:H06 | ES 151 |
| M00042584B:C10 | ES 151 |
| M00042623D:D07 | ES 151 |
| M00042625C:B04 | ES 151 |
| M00042626B:D08 | ES 151 |
| M00042627C:D01 | ES 151 |
| M00042630A:C05 | ES 151 |
| M00042955C:D05 | ES 151 |
| M00042956C:B06 | ES 151 |
| M00042960D:H08 | ES 151 |
| M00042962D:C05 | ES 151 |
| M00042964D:A03 | ES 151 |
| M00042966B:F07 | ES 151 |
| M00042966C:E06 | ES 151 |
| M00042970C:A04 | ES 151 |
| M00042970C:H10 | ES 151 |
| M00042976A:H04 | ES 151 |
| M00042979B:E02 | ES 151 |
| M00042981B:D11 | ES 151 |
| M00042983C:A11 | ES 151 |
| M00042983C:G06 | ES 151 |
| M00042986C:G12 | ES 151 |
| M00042988A:F06 | ES 151 |
| M00042997B:D06 | ES 151 |
| M00042998A:E03 | ES 151 |
| M00042998A:G04 | ES 151 |
| M00043001B:H10 | ES 151 |
| M00043001D:D03 | ES 151 |
| M00043002A:E05 | ES 151 |
| M00043003C:D08 | ES 151 |
| M00043011A:H12 | ES 151 |
| M00043015A:H10 | ES 151 |
| M00043022A:E12 | ES 151 |
| M00043026C:D07 | ES 151 |
| M00043028A:G05 | ES 151 |
| M00043029C:A06 | ES 151 |
| M00043032C:A10 | ES 151 |
| M00043034D:C01 | ES 151 |
| M00043036C:E05 | ES 151 |
| M00043036D:C09 | ES 151 |
| M00043040B:B07 | ES 151 |
| M00043044B:A12 | ES 151 |
| M00043044D:A09 | ES 151 |
| M00043045D:G12 | ES 151 |

| Clone Name     | Tube   |
|----------------|--------|
| M00043046D:B11 | ES 151 |
| M00043060D:G12 | ES 151 |
| M00043066B:H11 | ES 151 |
| M00043067D:D10 | ES 151 |
| M00043125A:B11 | ES 151 |
| M00043125C:A11 | ES 151 |
| M00042611A:A06 | ES 151 |
| M00042611D:B12 | ES 151 |
| M00042612D:F06 | ES 151 |
| M00042614B:B05 | ES 151 |
| M00043073A:C12 | ES 151 |
| M00043078D:D04 | ES 151 |
| M00043081D:F05 | ES 151 |
| M00043087B:G07 | ES 151 |
| M00043093C:G11 | ES 151 |
| M00043095A:F09 | ES 152 |
| M00043096A:G04 | ES 152 |
| M00043108A:F06 | ES 152 |
| M00043109C:G01 | ES 152 |
| M00043131B:A09 | ES 152 |
| M00043133B:C11 | ES 152 |
| M00043138D:B11 | ES 152 |
| M00043143B:A10 | ES 152 |
| M00043148C:A09 | ES 152 |
| M00043154A:B07 | ES 152 |
| M00043162A:B08 | ES 152 |
| M00043162D:C12 | ES 152 |
| M00043164C:E12 | ES 152 |
| M00043165B:G01 | ES 152 |
| M00043173D:G03 | ES 152 |
| M00043184A:H08 | ES 152 |
| M00043187A:C04 | ES 152 |
| M00043191A:A07 | ES 152 |
| M00043192C:B12 | ES 152 |
| M00043200A:H09 | ES 152 |
| M00043200B:C08 | ES 152 |
| M00043202B:F01 | ES 152 |
| M00043203A:B09 | ES 152 |
| M00043210C:E05 | ES 152 |
| M00043211A:F01 | ES 152 |
| M00043213B:B12 | ES 152 |
| M00043215A:D02 | ES 152 |
| M00043220B:C04 | ES 152 |
| M00042591D:H03 | ES 152 |

| Clone Name     | Tube   |
|----------------|--------|
| M00042592A:H10 | ES 152 |
| M00042593A:C02 | ES 152 |
| M00042593C:G06 | ES 152 |
| M00042595A:A11 | ES 152 |
| M00042595A:B01 | ES 152 |
| M00042596B:F06 | ES 152 |
| M00042596C:D07 | ES 152 |
| M00042597B:E12 | ES 152 |
| M00043416C:A02 | ES 152 |
| M00043417C:D05 | ES 152 |
| M00043418A:H10 | ES 152 |
| M00043419D:A10 | ES 152 |
| M00043428D:G08 | ES 152 |
| M00043430B:C02 | ES 152 |
| M00043431D:B08 | ES 152 |
| M00043433B:G09 | ES 152 |
| M00043433C:G07 | ES 152 |
| M00043437D:D04 | ES 152 |
| M00043440C:B07 | ES 152 |
| M00043446C:E12 | ES 152 |
| M00043447A:C07 | ES 152 |
| M00043449A:E12 | ES 152 |
| M00043450C:C06 | ES 152 |
| M00043453B:B09 | ES 152 |
| M00043458A:B12 | ES 152 |
| M00043461D:C02 | ES 152 |
| M00043461D:E06 | ES 152 |
| M00043465B:H02 | ES 152 |
| M00043465C:A03 | ES 152 |
| M00043465C:C09 | ES 152 |
| M00043476A:F07 | ES 152 |
| M00043483B:G10 | ES 152 |
| M00043491C:F04 | ES 152 |
| M00043492A:E01 | ES 152 |
| M00043513D:G08 | ES 152 |
| M00043516B:H09 | ES 152 |
| M00043518B:D06 | ES 152 |
| M00043526B:D10 | ES 152 |
| M00043527C:E09 | ES 152 |
| M00043528C:A02 | ES 152 |
| M00043616B:F02 | ES 152 |
| M00043616C:A05 | ES 152 |
| M00043632D:F09 | ES 152 |
| M00043634A:C10 | ES 152 |



| Clone Name     | Tube   |
|----------------|--------|
| M00043635C:C11 | ES 152 |
| M00043636B:C06 | ES 152 |
| M00043637C:H01 | ES 152 |
| M00043638A:D06 | ES 152 |
| M00043640C:E03 | ES 152 |
| M00043648A:G07 | ES 152 |
| M00043649B:E07 | ES 152 |
| M00001338C:B02 | ES 153 |
| M00001338C:F05 | ES 153 |
| M00001338D:D01 | ES 153 |
| M00001340D:F07 | ES 153 |
| M00001344D:E08 | ES 153 |
| M00001346B:G11 | ES 153 |
| M00001348B:B03 | ES 153 |
| M00001349C:B04 | ES 153 |
| M00001351B:E11 | ES 153 |
| M00001352B:B02 | ES 153 |
| M00001353A:H07 | ES 153 |
| M00001353C:A05 | ES 153 |
| M00001353D:E05 | ES 153 |
| M00001356D:E06 | ES 153 |
| M00001358A:E08 | ES 153 |
| M00001359A:H10 | ES 153 |
| M00001361A:C12 | ES 153 |
| M00001361B:A12 | ES 153 |
| M00001362A:F09 | ES 153 |
| M00001364A:C09 | ES 153 |
| M00001364C:H10 | ES 153 |
| M00001368A:A08 | ES 153 |
| M00001368A:B07 | ES 153 |
| M00001368A:C02 | ES 153 |
| M00001369A:G06 | ES 153 |
| M00001374A:B02 | ES 153 |
| M00001374C:B10 | ES 153 |
| M00001375B:D04 | ES 153 |
| M00001378C:E10 | ES 153 |
| M00001379A:F09 | ES 153 |
| M00001382D:A07 | ES 153 |
| M00001382D:H08 | ES 153 |
| M00001384A:A07 | ES 153 |
| M00001385A:E07 | ES 153 |
| M00001386B:F11 | ES 153 |
| M00001387A:C12 | ES 153 |
| M00001387B:A11 | ES 153 |
| M00001389B:E10 | ES 153 |
| M00001389D:D06 | ES 153 |
| M00001390D:E02 | ES 153 |

| Clone Name     | Tube   |
|----------------|--------|
| M00001391D:D03 | ES 153 |
| M00001393B:C03 | ES 153 |
| M00001393C:E08 | ES 153 |
| M00001393C:F04 | ES 153 |
| M00001393D:E02 | ES 153 |
| M00001396B:B01 | ES 153 |
| M00001396B:B12 | ES 153 |
| M00001396D:H02 | ES 153 |
| M00001397C:H08 | ES 153 |
| M00001399B:B01 | ES 153 |
| M00001399C:A01 | ES 153 |
| M00001403C:B03 | ES 153 |
| M00001403D:C12 | ES 153 |
| M00001406B:H09 | ES 153 |
| M00001406D:F06 | ES 153 |
| M00001410A:G10 | ES 153 |
| M00001416B:A05 | ES 153 |
| M00001421B:E07 | ES 153 |
| M00001422B:D06 | ES 153 |
| M00001424B:H06 | ES 153 |
| M00001424D:D02 | ES 153 |
| M00001426C:F06 | ES 153 |
| M00001428B:C10 | ES 153 |
| M00001429B:G05 | ES 153 |
| M00001430B:C01 | ES 153 |
| M00001433B:E02 | ES 153 |
| M00001442A:F08 | ES 153 |
| M00001442C:G12 | ES 153 |
| M00001444B:E04 | ES 153 |
| M00001444C:D11 | ES 153 |
| M00001445B:F06 | ES 153 |
| M00001449B:H10 | ES 153 |
| M00001451C:E10 | ES 153 |
| M00001460C:E10 | ES 153 |
| M00001461D:B10 | ES 153 |
| M00001461D:C10 | ES 153 |
| M00001465C:A02 | ES 153 |
| M00001466B:F03 | ES 153 |
| M00001467C:D04 | ES 153 |
| M00001477D:G09 | ES 153 |
| M00001485C:F06 | ES 153 |
| M00001488C:A03 | ES 153 |
| M00001497C:F10 | ES 153 |
| M00001503B:H10 | ES 153 |
| M00001506B:D11 | ES 153 |
| M00001512D:F08 | ES 153 |
| M00001518B:D10 | ES 153 |

| Clone Name     | Tube   |
|----------------|--------|
| M00001528C:C03 | ES 153 |
| M00001532A:G08 | ES 153 |
| M00001533C:G11 | ES 153 |
| M00001533D:A01 | ES 153 |
| M00001534C:E07 | ES 153 |
| M00001535B:B10 | ES 153 |
| M00001535B:E02 | ES 153 |
| M00001537B:H10 | ES 153 |
| M00001538B:A07 | ES 153 |
| M00001539C:F12 | ES 154 |
| M00001542B:F09 | ES 154 |
| M00001543C:A08 | ES 154 |
| M00001544B:B05 | ES 154 |
| M00001544B:E06 | ES 154 |
| M00001546B:C11 | ES 154 |
| M00001548B:D06 | ES 154 |
| M00001550A:H06 | ES 154 |
| M00001550D:B11 | ES 154 |
| M00001551D:D01 | ES 154 |
| M00001551D:H09 | ES 154 |
| M00001554C:G10 | ES 154 |
| M00001558A:E06 | ES 154 |
| M00001559A:H09 | ES 154 |
| M00001561D:H04 | ES 154 |
| M00001562B:B02 | ES 154 |
| M00001562D:B07 | ES 154 |
| M00001565A:H05 | ES 154 |
| M00001568C:A03 | ES 154 |
| M00001570A:B07 | ES 154 |
| M00001591B:H05 | ES 154 |
| M00001596A:D02 | ES 154 |
| M00001600B:G01 | ES 154 |
| M00001605B:B05 | ES 154 |
| M00001606B:A10 | ES 154 |
| M00001606D:D06 | ES 154 |
| M00001607A:E04 | ES 154 |
| M00001607D:H09 | ES 154 |
| M00001609D:C11 | ES 154 |
| M00001616D:F03 | ES 154 |
| M00001617C:F10 | ES 154 |
| M00001618C:D01 | ES 154 |
| M00001619C:H09 | ES 154 |
| M00001620B:A03 | ES 154 |
| M00001623D:A10 | ES 154 |
| M00001623D:E12 | ES 154 |
| M00001624A:C01 | ES 154 |
| M00001625D:B04 | ES 154 |

| Clone Name     | Tube   |
|----------------|--------|
| M00001626A:D07 | ES 154 |
| M00001632C:A10 | ES 154 |
| M00001633D:C11 | ES 154 |
| M00001637D:C12 | ES 154 |
| M00001648A:D10 | ES 154 |
| M00001661D:F06 | ES 154 |
| M00001663A:A12 | ES 154 |
| M00001671A:H10 | ES 154 |
| M00001671C:F03 | ES 154 |
| M00001675B:D06 | ES 154 |
| M00001677B:H08 | ES 154 |
| M00001680A:A01 | ES 154 |
| M00001683B:F11 | ES 154 |
| M00001684D:E04 | ES 154 |
| M00001686B:H01 | ES 154 |
| M00001686D:F06 | ES 154 |
| M00001688B:B11 | ES 154 |
| M00001692C:C04 | ES 154 |
| M00001771B:E06 | ES 154 |
| M00003746C:E11 | ES 154 |
| M00003749C:C08 | ES 154 |
| M00003753A:C11 | ES 154 |
| M00003758B:D07 | ES 154 |
| M00003758B:F06 | ES 154 |
| M00003760C:G10 | ES 154 |
| M00003761B:B02 | ES 154 |
| M00003763A:B02 | ES 154 |
| M00003763B:B10 | ES 154 |
| M00003764A:H09 | ES 154 |
| M00003764B:F11 | ES 154 |
| M00003764B:H11 | ES 154 |
| M00003764D:F07 | ES 154 |
| M00003768D:D08 | ES 154 |
| M00003770C:A10 | ES 154 |
| M00003771D:A03 | ES 154 |
| M00003773A:F10 | ES 154 |
| M00003780A:G01 | ES 154 |
| M00003782A:B02 | ES 154 |
| M00003785D:F07 | ES 154 |
| M00003787D:A10 | ES 154 |
| M00003808A:F11 | ES 154 |
| M00003808B:E07 | ES 154 |
| M00003812C:A03 | ES 154 |
| M00003814A:G05 | ES 154 |
| M00003819B:B01 | ES 154 |
| M00003820B:F11 | ES 154 |
| M00003821C:E12 | ES 154 |

| Clone Name     | Tube   |
|----------------|--------|
| M00003822C:A09 | ES 154 |
| M00003822D:A02 | ES 154 |
| M00003823B:A06 | ES 154 |
| M00003825A:H10 | ES 154 |
| M00003828A:D11 | ES 154 |
| M00003830B:C06 | ES 154 |
| M00003830C:D02 | ES 154 |
| M00003837C:D10 | ES 154 |
| M00003839C:H10 | ES 154 |
| M00003842D:D11 | ES 154 |
| M00003842D:H09 | ES 154 |
| M00003845A:C07 | ES 155 |
| M00003845D:G03 | ES 155 |
| M00003847A:H04 | ES 155 |
| M00003848C:G09 | ES 155 |
| M00003851B:A01 | ES 155 |
| M00003854B:F07 | ES 155 |
| M00003855C:F02 | ES 155 |
| M00003884A:E12 | ES 155 |
| M00003887C:E09 | ES 155 |
| M00003888B:F09 | ES 155 |
| M00003891B:H02 | ES 155 |
| M00003898C:A01 | ES 155 |
| M00003900C:D12 | ES 155 |
| M00003906A:C02 | ES 155 |
| M00003911C:A09 | ES 155 |
| M00003914A:A08 | ES 155 |
| M00003915C:D10 | ES 155 |
| M00003915C:G08 | ES 155 |
| M00003916A:E04 | ES 155 |
| M00003926A:F11 | ES 155 |
| M00003935B:B01 | ES 155 |
| M00003938C:A05 | ES 155 |
| M00003942A:D01 | ES 155 |
| M00003958C:H08 | ES 155 |
| M00003959D:A05 | ES 155 |
| M00003960D:C12 | ES 155 |
| M00003963D:F01 | ES 155 |
| M00003965D:D11 | ES 155 |
| M00003968C:G03 | ES 155 |
| M00003970D:H07 | ES 155 |
| M00003972C:F07 | ES 155 |
| M00003974C:E11 | ES 155 |
| M00003974D:E02 | ES 155 |
| M00003979B:A04 | ES 155 |
| M00003980D:C06 | ES 155 |
| M00003985D:B02 | ES 155 |

| Clone Name     | Tube   |
|----------------|--------|
| M00003988D:B01 | ES 155 |
| M00003991A:C11 | ES 155 |
| M00003993C:D07 | ES 155 |
| M00003993D:B03 | ES 155 |
| M00003994A:B10 | ES 155 |
| M00003996B:H07 | ES 155 |
| M00003998B:G10 | ES 155 |
| M00004028B:F10 | ES 155 |
| M00004029D:A01 | ES 155 |
| M00004031C:G06 | ES 155 |
| M00004036B:A11 | ES 155 |
| M00004036D:C12 | ES 155 |
| M00004038A:A04 | ES 155 |
| M00004042B:A11 | ES 155 |
| M00004047C:B09 | ES 155 |
| M00004047D:F12 | ES 155 |
| M00004053D:F09 | ES 155 |
| M00004054A:D03 | ES 155 |
| M00004055C:B10 | ES 155 |
| M00004055D:D05 | ES 155 |
| M00004057D:G01 | ES 155 |
| M00004061B:E05 | ES 155 |
| M00004062D:A02 | ES 155 |
| M00004066D:G10 | ES 155 |
| M00004067B:D03 | ES 155 |
| M00004080C:C04 | ES 155 |
| M00004085A:H01 | ES 155 |
| M00004085B:H02 | ES 155 |
| M00004087C:E02 | ES 155 |
| M00004093A:C03 | ES 155 |
| M00004096D:F02 | ES 155 |
| M00004102A:E03 | ES 155 |
| M00004103C:E10 | ES 155 |
| M00004104A:A12 | ES 155 |
| M00004110D:F09 | ES 155 |
| M00004114C:D11 | ES 155 |
| M00004115A:G12 | ES 155 |
| M00004118C:D12 | ES 155 |
| M00004122C:D01 | ES 155 |
| M00004134A:A08 | ES 155 |
| M00004136C:B12 | ES 155 |
| M00004139B:F01 | ES 155 |
| M00004141A:D01 | ES 155 |
| M00004141B:B01 | ES 155 |
| M00004141B:F08 | ES 155 |
| M00004143B:B04 | ES 155 |
| M00004144D:B02 | ES 155 |

| Clone Name     | Tube   |
|----------------|--------|
| M00004146A:C11 | ES 155 |
| M00004146B:E08 | ES 155 |
| M00004146C:B04 | ES 155 |
| M00004147C:E01 | ES 155 |
| M00004151B:A07 | ES 155 |
| M00004155A:H03 | ES 155 |
| M00004155C:A10 | ES 155 |
| M00004158B:E03 | ES 155 |
| M00004158D:E08 | ES 155 |
| M00004159C:D10 | ES 155 |
| M00004159D:F12 | ES 155 |
| M00004160D:F06 | ES 155 |
| M00004160D:G05 | ES 155 |
| M00004162D:F02 | ES 156 |
| M00004163B:C03 | ES 156 |
| M00004163C:A03 | ES 156 |
| M00004164B:E12 | ES 156 |
| M00004165C:A11 | ES 156 |
| M00004166C:B10 | ES 156 |
| M00004169A:E04 | ES 156 |
| M00004170A:F03 | ES 156 |
| M00004171B:B03 | ES 156 |
| M00004172C:A08 | ES 156 |
| M00004172D:B12 | ES 156 |
| M00004172D:F04 | ES 156 |
| M00004175D:E06 | ES 156 |
| M00004176C:A09 | ES 156 |
| M00004179C:B06 | ES 156 |
| M00004179D:A12 | ES 156 |
| M00004187B:C02 | ES 156 |
| M00004189A:C12 | ES 156 |
| M00004192C:B06 | ES 156 |
| M00004195A:F07 | ES 156 |
| M00004200C:A04 | ES 156 |
| M00004201D:C01 | ES 156 |
| M00004201D:C03 | ES 156 |
| M00004204C:H08 | ES 156 |
| M00004207C:A04 | ES 156 |
| M00004208A:D08 | ES 156 |
| M00004210A:A03 | ES 156 |
| M00004212D:C03 | ES 156 |
| M00004214A:E05 | ES 156 |
| M00004214D:A05 | ES 156 |
| M00004215B:C05 | ES 156 |
| M00004220D:C11 | ES 156 |
| M00004225D:E03 | ES 156 |
| M00004229B:B06 | ES 156 |

| Clone Name     | Tube   |
|----------------|--------|
| M00004230D:B05 | ES 156 |
| M00004237C:D10 | ES 156 |
| M00004242D:H01 | ES 156 |
| M00004245C:G10 | ES 156 |
| M00004246B:H07 | ES 156 |
| M00004251D:D03 | ES 156 |
| M00004263C:D03 | ES 156 |
| M00004266B:F07 | ES 156 |
| M00004269A:F11 | ES 156 |
| M00004269A:G11 | ES 156 |
| M00004269B:B04 | ES 156 |
| M00004270A:E09 | ES 156 |
| M00004276C:A08 | ES 156 |
| M00004277D:B02 | ES 156 |
| M00004278A:G06 | ES 156 |
| M00004278C:B10 | ES 156 |
| M00004281A:C04 | ES 156 |
| M00004282A:D01 | ES 156 |
| M00004282B:D07 | ES 156 |
| M00004282C:A12 | ES 156 |
| M00004284A:F08 | ES 156 |
| M00004295D:C07 | ES 156 |
| M00004296B:D03 | ES 156 |
| M00004303C:C05 | ES 156 |
| M00004310B:E02 | ES 156 |
| M00004316A:B03 | ES 156 |
| M00004320C:E07 | ES 156 |
| M00004321C:C11 | ES 156 |
| M00004322B:D03 | ES 156 |
| M00004324A:B03 | ES 156 |
| M00004324A:D10 | ES 156 |
| M00004324A:D05 | ES 156 |
| M00004328A:D01 | ES 156 |
| M00004330A:A01 | ES 156 |
| M00004336A:A01 | ES 156 |
| M00004341C:A09 | ES 156 |
| M00004341C:E05 | ES 156 |
| M00004344A:G11 | ES 156 |
| M00004344D:C12 | ES 156 |
| M00004347B:E04 | ES 156 |
| M00004347C:A05 | ES 156 |
| M00004350A:A04 | ES 156 |
| M00004351B:G07 | ES 156 |
| M00004352A:D08 | ES 156 |
| M00004357B:B06 | ES 156 |
| M00004358B:G02 | ES 156 |
| M00004359A:E01 | ES 156 |

| Clone Name     | Tube   |
|----------------|--------|
| M00004360C:D09 | ES 156 |
| M00004365C:C09 | ES 156 |
| M00004365C:G11 | ES 156 |
| M00004366D:C11 | ES 156 |
| M00004368A:B11 | ES 156 |
| M00004372A:E12 | ES 156 |
| M00004376D:A12 | ES 156 |
| M00004385C:H12 | ES 156 |
| M00004393C:D06 | ES 156 |
| M00004406A:G09 | ES 156 |
| M00004416B:G10 | ES 156 |
| M00004418B:A11 | ES 156 |
| M00004419A:G02 | ES 156 |
| M00004420D:E05 | ES 156 |
| M00004430A:A05 | ES 156 |
| M00004430B:B10 | ES 157 |
| M00004443C:F07 | ES 157 |
| M00004462D:D12 | ES 157 |
| M00004502A:D12 | ES 157 |
| M00004507D:E03 | ES 157 |
| M00004509B:B10 | ES 157 |
| M00004509D:C06 | ES 157 |
| M00004603B:E02 | ES 157 |
| M00004603C:C10 | ES 157 |
| M00004606D:H09 | ES 157 |
| M00004608A:C10 | ES 157 |
| M00004608A:H04 | ES 157 |
| M00004609A:E09 | ES 157 |
| M00023389A:G04 | ES 157 |
| M00023394D:D10 | ES 157 |
| M00026809A:H08 | ES 157 |
| M00026818C:E01 | ES 157 |
| M00026836B:H03 | ES 157 |
| M00026842B:A01 | ES 157 |
| M00026842D:C02 | ES 157 |
| M00026850B:C09 | ES 157 |
| M00026856B:G03 | ES 157 |
| M00026900A:H07 | ES 157 |
| M00026907D:E07 | ES 157 |
| M00026910B:G06 | ES 157 |
| M00026914C:H09 | ES 157 |
| M00026936D:C07 | ES 157 |
| M00026961A:B06 | ES 157 |
| M00026994D:D07 | ES 157 |
| M00027004C:C11 | ES 157 |
| M00027017A:B09 | ES 157 |
| M00027036A:B06 | ES 157 |

| Clone Name     | Tube   |
|----------------|--------|
| M00027050A:B02 | ES 157 |
| M00027052A:E10 | ES 157 |
| M00027057C:D10 | ES 157 |
| M00027064B:D06 | ES 157 |
| M00027081A:A08 | ES 157 |
| M00027093A:H02 | ES 157 |
| M00027131A:B03 | ES 157 |
| M00027159C:F07 | ES 157 |
| M00027167C:B10 | ES 157 |
| M00027168B:H08 | ES 157 |
| M00027170D:C07 | ES 157 |
| M00027173C:E11 | ES 157 |
| M00027177B:D04 | ES 157 |
| M00027178B:A11 | ES 157 |
| M00027182B:G06 | ES 157 |
| M00027189C:B10 | ES 157 |
| M00027193C:A07 | ES 157 |
| M00027215A:F06 | ES 157 |
| M00027215B:B12 | ES 157 |
| M00027244C:B06 | ES 157 |
| M00027247C:D02 | ES 157 |
| M00027262A:A07 | ES 157 |
| M00027270A:D04 | ES 157 |
| M00027274A:A09 | ES 157 |
| M00027290C:F06 | ES 157 |
| M00027291A:G08 | ES 157 |
| M00027311A:H09 | ES 157 |
| M00027313C:E01 | ES 157 |
| M00027314D:E02 | ES 157 |
| M00027316C:C03 | ES 157 |
| M00027319C:C03 | ES 157 |
| M00027319D:F07 | ES 157 |
| M00027332B:H09 | ES 157 |
| M00027359B:A06 | ES 157 |
| M00027363D:G04 | ES 157 |
| M00027364B:E12 | ES 157 |
| M00027376C:A02 | ES 157 |
| M00027381B:B04 | ES 157 |
| M00027400D:H02 | ES 157 |
| M00027433B:D12 | ES 157 |
| M00027457B:E11 | ES 157 |
| M00027459C:B10 | ES 157 |
| M00027467A:C07 | ES 157 |
| M00027475D:A01 | ES 157 |
| M00027480C:E09 | ES 157 |
| M00027485C:F07 | ES 157 |
| M00027506B:G01 | ES 157 |

| Clone Name     | Tube   |
|----------------|--------|
| M00027513D:F06 | ES 157 |
| M00027523A:H05 | ES 157 |
| M00027527B:C05 | ES 157 |
| M00027549C:G03 | ES 157 |
| M00027569A:E05 | ES 157 |
| M00027586A:C09 | ES 157 |
| M00027589B:G07 | ES 157 |
| M00027591A:E04 | ES 157 |
| M00027600B:C07 | ES 157 |
| M00027605B:D09 | ES 157 |
| M00027688C:C01 | ES 157 |
| M00027717C:C06 | ES 157 |
| M00027724D:D04 | ES 157 |
| M00027734D:C03 | ES 157 |
| M00027746A:D06 | ES 157 |
| M00027801B:D07 | ES 157 |
| M00027806C:H05 | ES 157 |
| M00028055B:G07 | ES 158 |
| M00028063C:H01 | ES 158 |
| M00028067A:C11 | ES 158 |
| M00028069D:H02 | ES 158 |
| M00028070A:H09 | ES 158 |
| M00028070D:C03 | ES 158 |
| M00028188C:H11 | ES 158 |
| M00028193B:E07 | ES 158 |
| M00028196A:G03 | ES 158 |
| M00028210B:H03 | ES 158 |
| M00028211A:F10 | ES 158 |
| M00028212D:C05 | ES 158 |
| M00028219B:H05 | ES 158 |
| M00028361B:H08 | ES 158 |
| M00028366B:B08 | ES 158 |
| M00028616C:D09 | ES 158 |
| M00028620C:C07 | ES 158 |
| M00028763A:G11 | ES 158 |
| M00028764B:D03 | ES 158 |
| M00028771A:E02 | ES 158 |
| M00028773C:C05 | ES 158 |
| M00028774D:E10 | ES 158 |
| M00028777B:G04 | ES 158 |
| M00028782A:F01 | ES 158 |
| M00028784A:D12 | ES 158 |
| M00028786B:A04 | ES 158 |
| M00031370B:C01 | ES 158 |
| M00031416D:H05 | ES 158 |

| Clone Name     | Tube   |
|----------------|--------|
| M00031484A:D03 | ES 158 |
| M00031485B:G05 | ES 158 |
| M00032471D:A05 | ES 158 |
| M00032473B:A03 | ES 158 |
| M00032474A:G03 | ES 158 |
| M00032475A:A06 | ES 158 |
| M00032489B:G12 | ES 158 |
| M00032490D:E08 | ES 158 |
| M00032494C:H08 | ES 158 |
| M00032497D:B10 | ES 158 |
| M00032504B:B10 | ES 158 |
| M00032507D:G08 | ES 158 |
| M00032508A:E03 | ES 158 |
| M00032515A:B12 | ES 158 |
| M00032517C:E10 | ES 158 |
| M00032519D:F08 | ES 158 |
| M00032534B:E12 | ES 158 |
| M00032541C:G03 | ES 158 |
| M00032553A:A07 | ES 158 |
| M00032556D:A03 | ES 158 |
| M00032562C:F01 | ES 158 |
| M00032567B:G05 | ES 158 |
| M00032568B:F08 | ES 158 |
| M00032577D:F01 | ES 158 |
| M00032580D:A09 | ES 158 |
| M00032581B:A09 | ES 158 |
| M00032584A:D06 | ES 158 |
| M00032586C:B04 | ES 158 |
| M00032590B:H01 | ES 158 |
| M00032594C:F05 | ES 158 |
| M00032597A:H02 | ES 158 |
| M00032605B:D09 | ES 158 |
| M00032613A:E11 | ES 158 |
| M00032614C:B10 | ES 158 |
| M00032614D:D08 | ES 158 |
| M00032620B:F06 | ES 158 |
| M00032621A:F11 | ES 158 |
| M00032628C:B06 | ES 158 |
| M00032634B:D09 | ES 158 |
| M00032637A:F09 | ES 158 |
| M00032638B:F02 | ES 158 |
| M00032644C:B05 | ES 158 |
| M00032645D:C01 | ES 158 |
| M00032647B:F06 | ES 158 |

| Clone Name     | Tube   |
|----------------|--------|
| M00032652C:C07 | ES 158 |
| M00032666A:C02 | ES 158 |
| M00032671B:D06 | ES 158 |
| M00032671B:D08 | ES 158 |
| M00032676C:C10 | ES 158 |
| M00032688C:A03 | ES 158 |
| M00032700A:E09 | ES 158 |
| M00032707D:F08 | ES 158 |
| M00032711B:F01 | ES 158 |
| M00032723D:H02 | ES 158 |
| M00032727A:E04 | ES 158 |
| M00032728D:F01 | ES 158 |
| M00032729A:F10 | ES 158 |
| M00032733B:F12 | ES 158 |
| M00032734B:E12 | ES 158 |
| M00032734C:C05 | ES 158 |
| M00032749D:G03 | ES 158 |
| M00032753A:C07 | ES 158 |
| M00032759A:A03 | ES 158 |
| M00032765A:C05 | ES 158 |
| M00032770C:G11 | ES 158 |
| M00032772D:D03 | ES 158 |
| M00032773D:F08 | ES 158 |
| M00032774C:C04 | ES 158 |
| M00032787D:C05 | ES 159 |
| M00032791B:H11 | ES 159 |
| M00032791D:F01 | ES 159 |
| M00032792C:B01 | ES 159 |
| M00032793A:G06 | ES 159 |
| M00032795C:A03 | ES 159 |
| M00032797D:D08 | ES 159 |
| M00032825B:F08 | ES 159 |
| M00032826C:D10 | ES 159 |
| M00032828A:A06 | ES 159 |
| M00032829D:A05 | ES 159 |
| M00032830D:D02 | ES 159 |
| M00032831A:C07 | ES 159 |
| M00032831A:E09 | ES 159 |
| M00032835D:G04 | ES 159 |
| M00032836B:A07 | ES 159 |
| M00032848D:B10 | ES 159 |
| M00032892C:C12 | ES 159 |
| M00032908A:D08 | ES 159 |
| M00032918D:B04 | ES 159 |

| Clone Name     | Tube   |
|----------------|--------|
| M00032928C:D02 | ES 159 |
| M00032944A:B07 | ES 159 |
| M00032945D:B07 | ES 159 |
| M00032979D:C11 | ES 159 |
| M00032979D:H07 | ES 159 |
| M00032985D:G09 | ES 159 |
| M00032987B:F01 | ES 159 |
| M00032988B:G01 | ES 159 |
| M00033006A:F10 | ES 159 |
| M00033028C:A02 | ES 159 |
| M00033028D:C10 | ES 159 |
| M00033037B:F04 | ES 159 |
| M00033041A:B11 | ES 159 |
| M00033055D:D02 | ES 159 |
| M00033071C:G05 | ES 159 |
| M00033071D:E08 | ES 159 |
| M00033072A:A09 | ES 159 |
| M00033080C:A07 | ES 159 |
| M00033081D:D11 | ES 159 |
| M00033144A:D02 | ES 159 |
| M00033146D:A03 | ES 159 |
| M00033147C:B08 | ES 159 |
| M00033149B:E10 | ES 159 |
| M00033150B:E02 | ES 159 |
| M00033150C:A11 | ES 159 |
| M00033183B:F10 | ES 159 |
| M00033218C:F07 | ES 159 |
| M00033223C:G04 | ES 159 |
| M00033230C:G10 | ES 159 |
| M00033232B:C08 | ES 159 |
| M00033246A:H12 | ES 159 |
| M00033248D:H11 | ES 159 |
| M00033264B:E06 | ES 159 |
| M00033274D:F03 | ES 159 |
| M00033311B:G10 | ES 159 |
| M00033324B:F04 | ES 159 |
| M00033326B:B05 | ES 159 |
| M00033329C:C02 | ES 159 |
| M00033296C:C11 | ES 159 |
| M00033302A:E11 | ES 159 |
| M00033302B:F10 | ES 159 |
| M00033303C:F09 | ES 159 |
| M00033342B:F03 | ES 159 |
| M00033344A:B06 | ES 159 |

| Clone Name     | Tube   |
|----------------|--------|
| M00033359C:H05 | ES 159 |
| M00033360C:A03 | ES 159 |
| M00033374D:C07 | ES 159 |
| M00033413A:A08 | ES 159 |
| M00033420B:E08 | ES 159 |
| M00033434D:F05 | ES 159 |
| M00033441A:B12 | ES 159 |
| M00033445D:G03 | ES 159 |
| M00038290A:D12 | ES 159 |
| M00038304B:E02 | ES 159 |
| M00038389D:D10 | ES 159 |
| M00038390B:F02 | ES 159 |
| M00038616C:C09 | ES 159 |
| M00038616D:B07 | ES 159 |
| M00038618D:D08 | ES 159 |
| M00038619B:F09 | ES 159 |
| M00038619D:C12 | ES 159 |
| M00039001A:B10 | ES 159 |
| M00039024D:E12 | ES 159 |
| M00039055C:A01 | ES 159 |
| M00039056B:G01 | ES 159 |
| M00039063C:H09 | ES 159 |
| M00039067A:C05 | ES 159 |
| M00039067B:F07 | ES 159 |
| M00039076D:G04 | ES 159 |
| M00039078B:B03 | ES 159 |
| M00039078D:C10 | ES 159 |
| M00039081B:C04 | ES 159 |
| M00039081B:G07 | ES 159 |
| M00039100A:G04 | ES 159 |
| M00039105D:A08 | ES 159 |
| M00039107A:E12 | ES 159 |
| M00039111A:C12 | ES 160 |
| M00039121D:E07 | ES 160 |
| M00039124D:H01 | ES 160 |
| M00039125D:H12 | ES 160 |
| M00039131C:B09 | ES 160 |
| M00039133B:D06 | ES 160 |
| M00039133C:F12 | ES 160 |
| M00039134D:F08 | ES 160 |
| M00039138B:G05 | ES 160 |
| M00039140A:F05 | ES 160 |
| M00039143A:F04 | ES 160 |
| M00039143D:C10 | ES 160 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039146B:G04 | ES 160 |
| M00039162D:C04 | ES 160 |
| M00039165D:C04 | ES 160 |
| M00039175A:F01 | ES 160 |
| M00039204A:E09 | ES 160 |
| M00039207A:F07 | ES 160 |
| M00039219B:C08 | ES 160 |
| M00039222B:A04 | ES 160 |
| M00039225A:D11 | ES 160 |
| M00039246B:A08 | ES 160 |
| M00039248C:A08 | ES 160 |
| M00039251C:H12 | ES 160 |
| M00039251D:B08 | ES 160 |
| M00039255D:B01 | ES 160 |
| M00039258C:C01 | ES 160 |
| M00039270D:D02 | ES 160 |
| M00039275B:E02 | ES 160 |
| M00039278C:D03 | ES 160 |
| M00039284D:H07 | ES 160 |
| M00039285B:G04 | ES 160 |
| M00039291D:F02 | ES 160 |
| M00039294C:B09 | ES 160 |
| M00039302B:E10 | ES 160 |
| M00039326A:G07 | ES 160 |
| M00039326C:B08 | ES 160 |
| M00039331B:F09 | ES 160 |
| M00039338B:F07 | ES 160 |
| M00039344C:A11 | ES 160 |
| M00039349D:B11 | ES 160 |
| M00039381C:C07 | ES 160 |
| M00039383A:H07 | ES 160 |
| M00039411D:D09 | ES 160 |
| M00039413C:E06 | ES 160 |
| M00039430A:E04 | ES 160 |
| M00039433B:D06 | ES 160 |
| M00039433C:E03 | ES 160 |
| M00039438B:D08 | ES 160 |
| M00039440C:G06 | ES 160 |
| M00039457D:C02 | ES 160 |
| M00039471D:G10 | ES 160 |
| M00039472B:E05 | ES 160 |
| M00039478C:B02 | ES 160 |
| M00039554D:B09 | ES 160 |
| M00039556C:G05 | ES 160 |



| Clone Name     | Tube   |
|----------------|--------|
| M00039559B:C07 | ES 160 |
| M00039560B:G09 | ES 160 |
| M00039560C:G06 | ES 160 |
| M00039617C:A10 | ES 160 |
| M00039654C:C11 | ES 160 |
| M00039668C:F01 | ES 160 |
| M00039672C:D05 | ES 160 |
| M00039686C:C01 | ES 160 |
| M00039694C:H01 | ES 160 |
| M00039698C:B03 | ES 160 |
| M00039710B:A01 | ES 160 |
| M00039710B:E01 | ES 160 |
| M00039785C:H12 | ES 160 |
| M00039786D:A10 | ES 160 |
| M00039805B:B06 | ES 160 |
| M00039806B:D05 | ES 160 |
| M00039820B:F06 | ES 160 |
| M00039822A:H02 | ES 160 |
| M00039826B:F09 | ES 160 |
| M00039826D:E04 | ES 160 |
| M00039828B:H06 | ES 160 |
| M00039829B:E01 | ES 160 |
| M00039860B:E01 | ES 160 |
| M00039860D:B02 | ES 160 |
| M00039861C:B12 | ES 160 |
| M00039865A:C09 | ES 160 |
| M00039869A:H01 | ES 160 |
| M00039871C:G05 | ES 160 |
| M00039873B:H04 | ES 160 |
| M00039874A:B06 | ES 160 |
| M00039885C:D11 | ES 160 |
| M00039894C:D09 | ES 160 |
| M00039895D:C04 | ES 160 |
| M00039900B:G04 | ES 160 |
| M00039915B:E08 | ES 160 |
| M00039921A:B10 | ES 160 |
| M00004824A:D12 | ES 160 |
| M00004824D:H05 | ES 160 |
| M00004831C:G11 | ES 160 |
| M00004832D:G04 | ES 160 |
| M00004836B:C02 | ES 161 |
| M00004839B:C12 | ES 161 |
| M00004843A:G12 | ES 161 |
| M00004846A:A10 | ES 161 |

| Clone Name     | Tube   |
|----------------|--------|
| M00004850A:B02 | ES 161 |
| M00004852D:C06 | ES 161 |
| M00004856D:F09 | ES 161 |
| M00004873B:G04 | ES 161 |
| M00004876B:A06 | ES 161 |
| M00005002A:C03 | ES 161 |
| M00005003D:C02 | ES 161 |
| M00005013D:H05 | ES 161 |
| M00005014B:F02 | ES 161 |
| M00005016C:E04 | ES 161 |
| M00005309B:A11 | ES 161 |
| M00005314A:G10 | ES 161 |
| M00005332A:C06 | ES 161 |
| M00005333D:D08 | ES 161 |
| M00005346D:A03 | ES 161 |
| M00005349C:C02 | ES 161 |
| M00005359B:B08 | ES 161 |
| M00005359B:D09 | ES 161 |
| M00005364B:E10 | ES 161 |
| M00005365A:F05 | ES 161 |
| M00005366D:F08 | ES 161 |
| M00005367D:A11 | ES 161 |
| M00005375D:A10 | ES 161 |
| M00005379A:D10 | ES 161 |
| M00005380B:H10 | ES 161 |
| M00005383A:C11 | ES 161 |
| M00005385A:B12 | ES 161 |
| M00005385D:F07 | ES 161 |
| M00005387A:B03 | ES 161 |
| M00005392A:G06 | ES 161 |
| M00005401D:F09 | ES 161 |
| M00005403C:A01 | ES 161 |
| M00005405C:D01 | ES 161 |
| M00005409D:B02 | ES 161 |
| M00005413D:A05 | ES 161 |
| M00005422B:B08 | ES 161 |
| M00005422D:H02 | ES 161 |
| M00005422D:H10 | ES 161 |
| M00005423A:C11 | ES 161 |
| M00005423C:A10 | ES 161 |
| M00005423C:D07 | ES 161 |
| M00005434A:C03 | ES 161 |
| M00005442A:B10 | ES 161 |
| M00005445A:E07 | ES 161 |

| Clone Name     | Tube   |
|----------------|--------|
| M00005445D:D04 | ES 161 |
| M00005445D:F11 | ES 161 |
| M00005452B:G03 | ES 161 |
| M00005452D:E05 | ES 161 |
| M00005460D:C11 | ES 161 |
| M00005461A:D12 | ES 161 |
| M00005463A:G02 | ES 161 |
| M00005466C:B01 | ES 161 |
| M00005468A:C04 | ES 161 |
| M00005468D:C01 | ES 161 |
| M00005474C:H09 | ES 161 |
| M00005485C:H04 | ES 161 |
| M00005489B:C08 | ES 161 |
| M00005500A:D04 | ES 161 |
| M00005504C:F12 | ES 161 |
| M00005504D:F06 | ES 161 |
| M00005505A:F01 | ES 161 |
| M00005505B:E01 | ES 161 |
| M00005506C:E09 | ES 161 |
| M00005506D:E11 | ES 161 |
| M00005507B:A03 | ES 161 |
| M00005511A:F05 | ES 161 |
| M00005512B:H01 | ES 161 |
| M00005515D:F02 | ES 161 |
| M00005520B:E01 | ES 161 |
| M00005520B:H05 | ES 161 |
| M00005524C:H04 | ES 161 |
| M00005535B:B01 | ES 161 |
| M00005540A:F09 | ES 161 |
| M00005557D:H10 | ES 161 |
| M00005569D:G09 | ES 161 |
| M00005570A:B08 | ES 161 |
| M00005570A:D05 | ES 161 |
| M00005603B:H03 | ES 161 |
| M00005606D:B12 | ES 161 |
| M00005607B:C04 | ES 161 |
| M00005616B:F07 | ES 161 |
| M00005622A:H02 | ES 161 |
| M00005623B:G01 | ES 161 |
| M00005626D:G11 | ES 161 |
| M00005634A:F07 | ES 161 |
| M00005641B:E09 | ES 161 |
| M00005643D:A05 | ES 161 |
| M00005674C:F04 | ES 161 |

| Clone Name     | Tube   |
|----------------|--------|
| M00005675D:D09 | ES 161 |
| M00005689C:B02 | ES 161 |
| M00005703B:E03 | ES 161 |
| M00005703D:G10 | ES 161 |
| M00005710B:H03 | ES 162 |
| M00005743D:A12 | ES 162 |
| M00005763D:A01 | ES 162 |
| M00005766D:D12 | ES 162 |
| M00005771D:C02 | ES 162 |
| M00005819D:F09 | ES 162 |
| M00005822C:A04 | ES 162 |
| M00006576D:C02 | ES 162 |
| M00006577A:H10 | ES 162 |
| M00006582D:A09 | ES 162 |
| M00006585A:D07 | ES 162 |
| M00006585A:F09 | ES 162 |
| M00006586D:D04 | ES 162 |
| M00006592A:A12 | ES 162 |
| M00006595B:C10 | ES 162 |
| M00006601D:G05 | ES 162 |
| M00006631C:A04 | ES 162 |
| M00006631D:D02 | ES 162 |
| M00006636B:E04 | ES 162 |
| M00006641B:F05 | ES 162 |
| M00006646A:A07 | ES 162 |
| M00006678A:A03 | ES 162 |
| M00006678C:C02 | ES 162 |
| M00006712C:H01 | ES 162 |
| M00006714C:D06 | ES 162 |
| M00006738A:F12 | ES 162 |
| M00006739B:A04 | ES 162 |
| M00006740B:A09 | ES 162 |
| M00006743A:D04 | ES 162 |
| M00006743A:H11 | ES 162 |
| M00006756B:G06 | ES 162 |
| M00006756C:A02 | ES 162 |
| M00006861D:H10 | ES 162 |
| M00006872D:B07 | ES 162 |
| M00006877B:C09 | ES 162 |
| M00006877C:F11 | ES 162 |
| M00006884D:A08 | ES 162 |
| M00006885A:F07 | ES 162 |
| M00006890C:F10 | ES 162 |
| M00006904D:A02 | ES 162 |

| Clone Name     | Tube   |
|----------------|--------|
| M00006907A:C09 | ES 162 |
| M00006907B:C06 | ES 162 |
| M00006989B:G05 | ES 162 |
| M00006994C:F06 | ES 162 |
| M00007002C:A10 | ES 162 |
| M00007006C:C12 | ES 162 |
| M00007007A:E04 | ES 162 |
| M00007031A:E02 | ES 162 |
| M00007032A:B05 | ES 162 |
| M00007032C:A12 | ES 162 |
| M00007046D:C09 | ES 162 |
| M00007048B:E11 | ES 162 |
| M00007048C:A12 | ES 162 |
| M00007059B:D07 | ES 162 |
| M00007060D:G07 | ES 162 |
| M00007064D:D12 | ES 162 |
| M00007070C:C01 | ES 162 |
| M00007081B:C08 | ES 162 |
| M00007081B:E09 | ES 162 |
| M00007082D:E05 | ES 162 |
| M00007098A:E10 | ES 162 |
| M00007103C:C12 | ES 162 |
| M00007103D:C02 | ES 162 |
| M00007112D:D03 | ES 162 |
| M00007117A:C11 | ES 162 |
| M00007126A:A02 | ES 162 |
| M00007141C:B05 | ES 162 |
| M00007154A:E06 | ES 162 |
| M00007155C:D07 | ES 162 |
| M00007155D:C09 | ES 162 |
| M00007158D:D03 | ES 162 |
| M00007178A:C02 | ES 162 |
| M00007195C:E11 | ES 162 |
| M00007197B:B05 | ES 162 |
| M00007202B:F01 | ES 162 |
| M00007947A:B06 | ES 162 |
| M00007953D:F07 | ES 162 |
| M00007969D:C01 | ES 162 |
| M00007973B:D11 | ES 162 |
| M00007975C:A10 | ES 162 |
| M00007975D:F12 | ES 162 |
| M00007980A:B01 | ES 162 |
| M00007980B:A07 | ES 162 |
| M00007981C:F07 | ES 162 |

| Clone Name     | Tube   |
|----------------|--------|
| M00007985C:D08 | ES 162 |
| M00008001B:F05 | ES 162 |
| M00008007B:E03 | ES 162 |
| M00008016B:E09 | ES 162 |
| M00008019B:A01 | ES 162 |
| M00008020D:D05 | ES 162 |
| M00008020D:F02 | ES 162 |
| M00008021C:G12 | ES 162 |
| M00008045C:A05 | ES 162 |
| M00008055D:G03 | ES 162 |
| M00008059B:F08 | ES 162 |
| M00008059D:B08 | ES 162 |
| M00008065D:A07 | ES 163 |
| M00008071D:H03 | ES 163 |
| M00008073A:D01 | ES 163 |
| M00008073D:D01 | ES 163 |
| M00021649B:A02 | ES 163 |
| M00021650D:A11 | ES 163 |
| M00021653A:B02 | ES 163 |
| M00021668D:A03 | ES 163 |
| M00021676C:G03 | ES 163 |
| M00021677A:D09 | ES 163 |
| M00021678A:H03 | ES 163 |
| M00021678D:H04 | ES 163 |
| M00021681C:C09 | ES 163 |
| M00021690A:C03 | ES 163 |
| M00021697C:B07 | ES 163 |
| M00021700D:H03 | ES 163 |
| M00021852C:H02 | ES 163 |
| M00021855D:F10 | ES 163 |
| M00021866C:H08 | ES 163 |
| M00021896D:A05 | ES 163 |
| M00021923A:B12 | ES 163 |
| M00021923D:H02 | ES 163 |
| M00021933B:F02 | ES 163 |
| M00021941A:D09 | ES 163 |
| M00021952B:G06 | ES 163 |
| M00021958B:E08 | ES 163 |
| M00021967D:H06 | ES 163 |
| M00021971C:B11 | ES 163 |
| M00021974D:F01 | ES 163 |
| M00021981A:C02 | ES 163 |
| M00021991D:F09 | ES 163 |
| M00021998B:D09 | ES 163 |

| Clone Name     | Tube   |
|----------------|--------|
| M00022009C:A08 | ES 163 |
| M00022016B:F01 | ES 163 |
| M00022032A:G05 | ES 163 |
| M00022051B:D07 | ES 163 |
| M00022069D:C12 | ES 163 |
| M00022070B:B04 | ES 163 |
| M00022073C:C07 | ES 163 |
| M00022081A:B07 | ES 163 |
| M00022088B:F10 | ES 163 |
| M00022088B:H02 | ES 163 |
| M00022088D:E10 | ES 163 |
| M00022090B:A10 | ES 163 |
| M00022092D:A11 | ES 163 |
| M00022094B:G02 | ES 163 |
| M00022096D:A03 | ES 163 |
| M00022103C:D05 | ES 163 |
| M00022104A:G08 | ES 163 |
| M00022117C:A02 | ES 163 |
| M00022118A:E06 | ES 163 |
| M00022140D:A07 | ES 163 |
| M00022144C:E12 | ES 163 |
| M00022158B:B09 | ES 163 |
| M00022170C:C01 | ES 163 |
| M00022171A:F03 | ES 163 |
| M00022185A:B03 | ES 163 |
| M00022193B:A09 | ES 163 |
| M00022193C:C09 | ES 163 |
| M00022200B:B05 | ES 163 |
| M00022202C:C04 | ES 163 |
| M00022208B:D03 | ES 163 |
| M00022208C:E04 | ES 163 |
| M00022208C:F08 | ES 163 |
| M00022212D:G02 | ES 163 |
| M00022216D:D10 | ES 163 |
| M00022218B:B12 | ES 163 |
| M00022220A:A07 | ES 163 |
| M00022224A:C07 | ES 163 |
| M00022224A:G07 | ES 163 |
| M00022228B:B11 | ES 163 |
| M00022229D:E01 | ES 163 |
| M00022237C:E04 | ES 163 |
| M00022237D:D06 | ES 163 |
| M00022238C:G04 | ES 163 |
| M00022240B:C12 | ES 163 |

| Clone Name     | Tube   |
|----------------|--------|
| M00022240D:B11 | ES 163 |
| M00022249D:C01 | ES 163 |
| M00022250A:B04 | ES 163 |
| M00022262A:F06 | ES 163 |
| M00022262B:B06 | ES 163 |
| M00022264A:B02 | ES 163 |
| M00022265A:F11 | ES 163 |
| M00022269C:A04 | ES 163 |
| M00022273A:E03 | ES 163 |
| M00022282B:C09 | ES 163 |
| M00022305A:B04 | ES 163 |
| M00022363C:D05 | ES 163 |
| M00022367D:G11 | ES 163 |
| M00022368A:B11 | ES 163 |
| M00022372D:H12 | ES 163 |
| M00022374C:E11 | ES 163 |
| M00022376D:D05 | ES 163 |
| M00022383C:A12 | ES 163 |
| M00022386D:F10 | ES 163 |
| M00022392B:F01 | ES 163 |
| M00022403C:E12 | ES 164 |
| M00022415C:D12 | ES 164 |
| M00022416D:D01 | ES 164 |
| M00022421A:F12 | ES 164 |
| M00022425A:C09 | ES 164 |
| M00022430C:C06 | ES 164 |
| M00022435B:G12 | ES 164 |
| M00022436C:F11 | ES 164 |
| M00022438C:H09 | ES 164 |
| M00022442B:G03 | ES 164 |
| M00022446C:H06 | ES 164 |
| M00022449D:F08 | ES 164 |
| M00022452B:E06 | ES 164 |
| M00022454C:B08 | ES 164 |
| M00022457A:G05 | ES 164 |
| M00022467D:B03 | ES 164 |
| M00022470D:B02 | ES 164 |
| M00022472D:B01 | ES 164 |
| M00022474B:C08 | ES 164 |
| M00022475D:C07 | ES 164 |
| M00022481B:A04 | ES 164 |
| M00022485B:E07 | ES 164 |
| M00022487B:A08 | ES 164 |
| M00022487C:C02 | ES 164 |

| Clone Name     | Tube   |
|----------------|--------|
| M00022491A:A08 | ES 164 |
| M00022491D:A10 | ES 164 |
| M00022494B:D06 | ES 164 |
| M00022494D:A05 | ES 164 |
| M00022499D:D08 | ES 164 |
| M00022507C:C08 | ES 164 |
| M00022509A:H02 | ES 164 |
| M00022509B:D11 | ES 164 |
| M00022512B:A09 | ES 164 |
| M00022516B:C05 | ES 164 |
| M00022525B:D09 | ES 164 |
| M00022530B:C04 | ES 164 |
| M00022537B:C06 | ES 164 |
| M00022546B:E05 | ES 164 |
| M00022559D:G10 | ES 164 |
| M00022563B:C08 | ES 164 |
| M00022590B:E05 | ES 164 |
| M00022600D:B05 | ES 164 |
| M00022601B:G06 | ES 164 |
| M00022618B:D09 | ES 164 |
| M00022618C:E04 | ES 164 |
| M00022627B:H03 | ES 164 |
| M00022634A:C07 | ES 164 |
| M00022634B:H09 | ES 164 |
| M00022638A:D03 | ES 164 |
| M00022642A:G08 | ES 164 |
| M00022648A:D08 | ES 164 |
| M00022656D:D07 | ES 164 |
| M00022662C:H04 | ES 164 |
| M00022662D:H03 | ES 164 |
| M00022672C:H04 | ES 164 |
| M00022674C:H08 | ES 164 |
| M00022677C:C01 | ES 164 |
| M00022678B:C08 | ES 164 |
| M00022681D:E10 | ES 164 |
| M00022682D:A10 | ES 164 |
| M00022684A:E06 | ES 164 |
| M00022690A:A07 | ES 164 |
| M00022694A:F05 | ES 164 |
| M00022696B:C11 | ES 164 |
| M00039921C:H11 | ES 164 |
| M00039929B:E06 | ES 164 |
| M00039929D:H10 | ES 164 |
| M00039932B:A07 | ES 164 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039976C:F11 | ES 164 |
| M00039977B:D12 | ES 164 |
| M00039981D:B01 | ES 164 |
| M00040003A:G10 | ES 164 |
| M00040016C:E07 | ES 164 |
| M00040023B:B10 | ES 164 |
| M00040025A:B04 | ES 164 |
| M00040034A:E06 | ES 164 |
| M00040034B:G02 | ES 164 |
| M00040041A:G08 | ES 164 |
| M00040041D:F01 | ES 164 |
| M00040045B:H07 | ES 164 |
| M00040061C:C08 | ES 164 |
| M00040075B:A05 | ES 164 |
| M00040078A:C07 | ES 164 |
| M00040079B:F06 | ES 164 |
| M00040079D:D09 | ES 164 |
| M00040081C:E02 | ES 164 |
| M00040094B:C08 | ES 164 |
| M00040118D:C05 | ES 164 |
| M00040123C:A10 | ES 164 |
| M00040127C:D02 | ES 164 |
| M00022698C:D10 | ES 164 |
| M00022702D:E02 | ES 164 |
| M00022703D:B11 | ES 164 |
| M00022706D:G08 | ES 164 |
| M00022727A:G01 | ES 164 |
| M00022738D:G08 | ES 164 |
| M00022740C:H11 | ES 165 |
| M00022797D:A06 | ES 165 |
| M00022801D:D09 | ES 165 |
| M00022805B:A10 | ES 165 |
| M00022812A:G01 | ES 165 |
| M00022820A:F07 | ES 165 |
| M00022835C:A09 | ES 165 |
| M00022854C:G07 | ES 165 |
| M00022856D:A07 | ES 165 |
| M00022857B:A09 | ES 165 |
| M00022897B:F06 | ES 165 |
| M00022901A:C05 | ES 165 |
| M00022904C:D04 | ES 165 |
| M00022924B:A05 | ES 165 |
| M00022924C:F04 | ES 165 |
| M00022945A:H09 | ES 165 |

| Clone Name     | Tube   |
|----------------|--------|
| M00022945B:F11 | ES 165 |
| M00022947B:D02 | ES 165 |
| M00022952A:B02 | ES 165 |
| M00022953B:D06 | ES 165 |
| M00022964A:B03 | ES 165 |
| M00022972C:E05 | ES 165 |
| M00022992A:H06 | ES 165 |
| M00022992B:G12 | ES 165 |
| M00022995C:G07 | ES 165 |
| M00023004C:A01 | ES 165 |
| M00023007D:D03 | ES 165 |
| M00023020C:H03 | ES 165 |
| M00023097D:B08 | ES 165 |
| M00039184D:H09 | ES 165 |
| M00039364D:E05 | ES 165 |
| M00039377B:E05 | ES 165 |
| M00039377B:H09 | ES 165 |
| M00039483A:D10 | ES 165 |
| M00039526A:A08 | ES 165 |
| M00039537A:F08 | ES 165 |
| M00039564D:D04 | ES 165 |
| M00039594C:B06 | ES 165 |
| M00039598A:E04 | ES 165 |
| M00039630D:B07 | ES 165 |
| M00039642A:A08 | ES 165 |
| M00039642C:F08 | ES 165 |
| M00039646A:E06 | ES 165 |
| M00039647A:A02 | ES 165 |
| M00039647B:A02 | ES 165 |
| M00039739B:H12 | ES 165 |
| M00040132A:H09 | ES 165 |
| M00040162A:E02 | ES 165 |
| M00040169A:G06 | ES 165 |
| M00040173D:A04 | ES 165 |
| M00040174D:G06 | ES 165 |
| M00040198A:F12 | ES 165 |
| M00040224C:F06 | ES 165 |
| M00040247D:D02 | ES 165 |
| M00040252C:G05 | ES 165 |
| M00040267D:A12 | ES 165 |
| M00040287A:C11 | ES 165 |
| M00040287C:F10 | ES 165 |
| M00040289D:C06 | ES 165 |
| M00039747B:B06 | ES 165 |

| Clone Name     | Tube   |
|----------------|--------|
| M00039748C:G09 | ES 165 |
| M00040201A:H01 | ES 165 |
| M00040219B:B07 | ES 165 |
| M00040291A:G10 | ES 165 |
| M00040298B:B09 | ES 165 |
| M00040314B:D07 | ES 165 |
| M00040326B:G09 | ES 165 |
| M00040329A:H05 | ES 165 |
| M00040338A:B10 | ES 165 |
| M00040344C:D05 | ES 165 |
| M00040349D:D07 | ES 165 |
| M00040351A:C08 | ES 165 |
| M00040351D:G07 | ES 165 |
| M00040366B:H10 | ES 165 |
| M00040367A:C08 | ES 165 |
| M00040381A:B06 | ES 165 |
| M00040384B:E04 | ES 165 |
| M00040391A:G05 | ES 165 |
| M00042525B:H01 | ES 165 |
| M00042528C:H01 | ES 165 |
| M00042554A:D01 | ES 165 |
| M00042557D:B06 | ES 165 |
| M00042560C:G06 | ES 165 |
| M00042579A:D09 | ES 165 |
| M00042719A:G08 | ES 165 |
| M00042722C:C09 | ES 165 |
| M00042724A:G06 | ES 165 |
| M00042732B:H06 | ES 165 |
| M00042734A:F05 | ES 165 |
| M00042742B:E04 | ES 165 |
| M00042743D:G10 | ES 165 |
| M00042891C:G08 | ES 165 |
| M00042894C:A11 | ES 165 |
| M00042908A:F09 | ES 165 |
| M00042915B:G11 | ES 165 |
| M00054793B:A06 | ES 165 |
| M00054911D:E06 | ES 166 |
| M00055430A:A01 | ES 166 |
| M00055433D:G03 | ES 166 |
| M00055448B:E05 | ES 166 |
| M00055454A:D02 | ES 166 |
| M00055456C:H06 | ES 166 |
| M00055466A:F06 | ES 166 |
| M00055468A:A08 | ES 166 |

| Clone Name     | Tube   |
|----------------|--------|
| M00055527B:E01 | ES 166 |
| M00055639A:E06 | ES 166 |
| M00055653C:B07 | ES 166 |
| M00055676A:G02 | ES 166 |
| M00055724B:E04 | ES 166 |
| M00055724D:C07 | ES 166 |
| M00055725D:D09 | ES 166 |
| M00055735A:H08 | ES 166 |
| M00055745B:A08 | ES 166 |
| M00055757A:B01 | ES 166 |
| M00055794A:E10 | ES 166 |
| M00055805A:H02 | ES 166 |
| M00055809A:B09 | ES 166 |
| M00055810C:D03 | ES 166 |
| M00055818B:D01 | ES 166 |
| M00055873D:C02 | ES 166 |
| M00055880B:H10 | ES 166 |
| M00055919B:C10 | ES 166 |
| M00055925D:B07 | ES 166 |
| M00055961C:B10 | ES 166 |
| M00055975B:F09 | ES 166 |
| M00055980C:B04 | ES 166 |
| M00056004B:C05 | ES 166 |
| M00056024B:F09 | ES 166 |
| M00056035D:A08 | ES 166 |
| M00056057C:F06 | ES 166 |
| M00056105A:D06 | ES 166 |
| M00056133A:E11 | ES 166 |
| M00056215D:F02 | ES 166 |
| M00056217D:E10 | ES 166 |
| M00056220D:G02 | ES 166 |
| M00056230D:E07 | ES 166 |
| M00056244A:B06 | ES 166 |
| M00056244C:H05 | ES 166 |
| M00056304A:H05 | ES 166 |
| M00056320B:A03 | ES 166 |
| M00056342A:C03 | ES 166 |
| M00056345D:A04 | ES 166 |
| M00056436C:F01 | ES 166 |
| M00056458C:E01 | ES 166 |
| M00042350A:A05 | ES 166 |
| M00042433A:E11 | ES 166 |
| M00042462B:C02 | ES 166 |
| M00042512D:D10 | ES 166 |

| Clone Name     | Tube   |
|----------------|--------|
| M00042766C:D05 | ES 166 |
| M00042788A:F04 | ES 166 |
| M00042794A:F01 | ES 166 |
| M00042796A:A10 | ES 166 |
| M00042801C:D01 | ES 166 |
| M00042822A:H04 | ES 166 |
| M00042857C:E01 | ES 166 |
| M00042858C:G11 | ES 166 |
| M00042860B:C07 | ES 166 |
| M00042863D:F09 | ES 166 |
| M00042878D:F05 | ES 166 |
| M00042878D:G06 | ES 166 |
| M00042352B:A04 | ES 166 |
| M00042352D:B03 | ES 166 |
| M00042449B:F05 | ES 166 |
| M00042457C:B06 | ES 166 |
| M00042516B:D01 | ES 166 |
| M00042520B:H04 | ES 166 |
| M00043299A:B10 | ES 166 |
| M00043306D:C01 | ES 166 |
| M00043313D:E09 | ES 166 |
| M00043328C:E04 | ES 166 |
| M00043336D:B03 | ES 166 |
| M00043339C:F11 | ES 166 |
| M00043355A:D07 | ES 166 |
| M00043358C:A02 | ES 166 |
| M00043402B:G07 | ES 166 |
| M00054499A:C08 | ES 166 |
| M00054528B:E05 | ES 166 |
| M00054536B:B01 | ES 166 |
| M00054538D:C12 | ES 166 |
| M00054542B:A10 | ES 166 |
| M00054548C:H06 | ES 166 |
| M00054569A:B07 | ES 166 |
| M00054579A:C02 | ES 166 |
| M00054599D:B03 | ES 166 |
| M00054623C:F05 | ES 166 |
| M00054643D:F07 | ES 166 |
| M00054675D:G03 | ES 166 |
| M00054682B:H02 | ES 166 |
| M00054683D:G11 | ES 166 |
| M00054686A:A09 | ES 166 |
| M00054686A:F10 | ES 166 |
| M00054693A:E11 | ES 166 |

| Clone Name     | Tube   |
|----------------|--------|
| M00054708C:B06 | ES 167 |
| M00054714B:G10 | ES 167 |
| M00054725C:D09 | ES 167 |
| M00054744C:F12 | ES 167 |
| M00054781B:H04 | ES 167 |
| M00054781D:A11 | ES 167 |
| M00054786C:D08 | ES 167 |
| M00054807D:C11 | ES 167 |
| M00054817D:A11 | ES 167 |
| M00054818B:F10 | ES 167 |
| M00054843A:C01 | ES 167 |
| M00054856C:D03 | ES 167 |
| M00054866B:C08 | ES 167 |
| M00054890C:D05 | ES 167 |
| M00054908C:A01 | ES 167 |
| M00054931D:E10 | ES 167 |
| M00054973B:E12 | ES 167 |
| M00054978C:F01 | ES 167 |
| M00055001C:G10 | ES 167 |
| M00055002B:E08 | ES 167 |
| M00055004C:H05 | ES 167 |
| M00055023A:E11 | ES 167 |
| M00055043B:H08 | ES 167 |
| M00055055C:F01 | ES 167 |
| M00055081A:A05 | ES 167 |
| M00055093B:A03 | ES 167 |
| M00055108B:A02 | ES 167 |
| M00055117A:E02 | ES 167 |
| M00055166C:D10 | ES 167 |
| M00055221C:H11 | ES 167 |
| M00055232A:E08 | ES 167 |
| M00055239D:F11 | ES 167 |
| M00055240A:A08 | ES 167 |
| M00055244B:F07 | ES 167 |
| M00055254A:H03 | ES 167 |
| M00055337B:C04 | ES 167 |
| M00055375C:F12 | ES 167 |
| M00055387C:C12 | ES 167 |
| M00055391B:C07 | ES 167 |
| M00055395D:D11 | ES 167 |
| M00055402A:H01 | ES 167 |
| M00055420A:E06 | ES 167 |
| M00055423A:B08 | ES 167 |
| M00055423C:G12 | ES 167 |

| Clone Name     | Tube   |
|----------------|--------|
| M00055423C:H10 | ES 167 |
| M00055424B:H06 | ES 167 |
| M00055424D:G05 | ES 167 |
| M00055425C:A04 | ES 167 |
| M00055473C:F02 | ES 167 |
| M00055477D:B01 | ES 167 |
| M00042585A:H11 | ES 167 |
| M00042585D:D03 | ES 167 |
| M00042585D:E10 | ES 167 |
| M00042586A:B01 | ES 167 |
| M00042588C:E02 | ES 167 |
| M00042621C:C04 | ES 167 |
| M00042951D:G12 | ES 167 |
| M00042960B:C06 | ES 167 |
| M00042967D:C01 | ES 167 |
| M00042970C:B01 | ES 167 |
| M00042972C:F04 | ES 167 |
| M00042976D:C01 | ES 167 |
| M00042982D:A10 | ES 167 |
| M00042986D:E03 | ES 167 |
| M00042996B:H08 | ES 167 |
| M00043013B:E03 | ES 167 |
| M00043015D:D05 | ES 167 |
| M00043016B:F09 | ES 167 |
| M00043017C:D08 | ES 167 |
| M00043063C:H05 | ES 167 |
| M00043070A:C03 | ES 167 |
| M00043113C:G09 | ES 167 |
| M00042617B:E01 | ES 167 |
| M00043074C:D07 | ES 167 |
| M00043076D:A02 | ES 167 |
| M00043077B:F11 | ES 167 |
| M00043077C:D12 | ES 167 |
| M00043077C:G10 | ES 167 |
| M00043099A:H04 | ES 167 |
| M00043101D:G11 | ES 167 |
| M00043134A:F05 | ES 167 |
| M00043152C:B10 | ES 167 |
| M00043213A:D05 | ES 167 |
| M00043219C:C02 | ES 167 |
| M00043221D:C12 | ES 167 |
| M00043222C:B06 | ES 167 |
| M00043455B:C08 | ES 167 |
| M00043465C:H11 | ES 167 |



| Clone Name     | Tube   |
|----------------|--------|
| M00043470A:C10 | ES 167 |
| M00043485C:C03 | ES 167 |
| M00043490C:F02 | ES 167 |
| M00043495C:H05 | ES 167 |

| Clone Name     | Tube   |
|----------------|--------|
| M00043528A:E11 | ES 167 |
| M00043529A:B08 | ES 167 |
| M00043640A:B01 | ES 167 |

TABLE 1

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 1      | 377044  | RTA00002676F.p.11.2.P.Seq | F           | M00039329A:C01 | CH09LNL   |
| 2      | 377708  | RTA00002683F.m.01.2.P.Seq | F           | M00040089A:G08 | CH09LNL   |
| 3      | 427782  | RTA00002666F.1.06.1.P.Seq | F           | M00032638D:A06 | CH08LNH   |
| 4      | 29372   | RTA00002712F.a.06.1.P.Seq | F           | M00023282A:C02 | CH04MAL   |
| 5      | 455003  | RTA00002694F.b.02.1.P.Seq | F           | M00043419D:A10 | CH20COHLV |
| 6      | 380625  | RTA00002684F.d.03.2.P.Seq | F           | M00040118D:G10 | CH09LNL   |
| 7      | 450959  | RTA00002691F.b.05.3.P.Seq | F           | M00043306D:B07 | CH17COHLV |
| 8      | 397851  | RTA00002680F.b.04.1.P.Seq | F           | M00039775A:A09 | CH09LNL   |
| 9      | 20652   | RTA00002710F.k.01.1.P.Seq | F           | M00022440B:E01 | CH03MAH   |
| 10     | 97930   | RTA00002663F.k.18.1.P.Seq | F           | M00022767B:G11 | CH03MAH   |
| 11     | 373071  | RTA00002670F.j.23.1.P.Seq | F           | M00033442A:D06 | CH09LNL   |
| 12     | 162369  | RTA00002713F.e.01.1.P.Seq | F           | M00027292D:F10 | CH04MAL   |
| 13     | 401247  | RTA00002685F.f.15.2.P.Seq | F           | M00039508A:C12 | CH12EDT   |
| 14     | 430738  | RTA00002669F.i.15.3.P.Seq | F           | M00033231D:B09 | CH08LNH   |
| 15     | 46779   | RTA00002711F.c.14.1P.Seq  | F           | M00022860C:G04 | CH3MAH    |
| 16     | 375772  | RTA00002681F.p.01.2.P.Seq | F           | M00039909C:G05 | CH09LNL   |
| 17     | 430689  | RTA00002669F.j.01.3.P.Seq | F           | M00033243B:A05 | CH08LNH   |
| 18     | 376546  | RTA00002677F.d.07.2.P.Seq | F           | M00039345C:C12 | CH09LNL   |
| 19     | 430041  | RTA00002667F.f.17.1.P.Seq | F           | M00032790B:A07 | CH08LNH   |
| 20     | 431643  | RTA00002669F.1.16.1.P.Seq | F           | M00033276D:H09 | CH08LNH   |
| 21     | 19422   | RTA00002709F.c.02.1.P.Seq | F           | M00005449B:B10 | CH02COH   |
| 22     | 376802  | RTA00002677F.c.18.2.P.Seq | F           | M00039344B:G07 | CH09LNL   |
| 23     | 376814  | RTA00002674F.h.02.1.P.Seq | F           | M00039139C:G12 | CH09LNL   |
| 24     | 375492  | RTA00002677F.m.19.2.P.Seq | F           | M00039418B:D08 | CH09LNL   |
| 25     | 379114  | RTA00002681F.n.24.2.P.Seq | F           | M00039903C:F03 | CH09LNL   |
| 26     | 380668  | RTA00002670F.p.11.1.P.Seq | F           | M00033581C:H10 | CH09LNL   |
| 27     | 213817  | RTA00002664F.i.19.2.P.Seq | F           | M00027634A:D11 | CH04MAL   |
| 28     | 375749  | RTA00002680F.f.23.1.P.Seq | F           | M00039795D:G06 | CH09LNL   |
| 29     | 430896  | RTA00002669F.b.20.4.P.Seq | F           | M00033185C:D01 | CH08LNH   |
| 30     | 380462  | RTA00002670F.o.01.1.P.Seq | F           | M00033570B:E06 | CH09LNL   |
| 31     | 430896  | RTA00002669F.b.20.3.P.Seq | F           | M00033185C:D01 | CH08LNH   |
| 32     | 376996  | RTA00002676F.p.13.2.P.Seq | F           | M00039329C:B10 | CH09LNL   |
| 33     | 374846  | RTA00002677F.k.19.2.P.Seq | F           | M00039412D:G06 | CH09LNL   |
| 34     | 379057  | RTA00002672F.n.13.2.P.Seq | F           | M00039039B:E03 | CH09LNL   |
| 35     | 374172  | RTA00002673F.k.16.2.P.Seq | F           | M00039097D:D06 | CH09LNL   |
| 36     | 373104  | RTA00002683F.o.15.2.P.Seq | F           | M00040098D:G12 | CH09LNL   |
| 37     | 186302  | RTA00002713F.m.21.1.P.Seq | F           | M00027591B:C04 | CH04MAL   |
| 38     | 427947  | RTA00002665F.o.01.1.P.Seq | F           | M00032495B:D02 | CH08LNH   |
| 39     | 375180  | RTA00002673F.d.17.1.P.Seq | F           | M00039064D:H09 | CH09LNL   |
| 40     | 377584  | RTA00002683F.1.22.2.P.Seq | F           | M00040088C:E10 | CH09LNL   |
| 41     | 377364  | RTA00002678F.a.15.2.P.Seq | F           | M00039432C:A01 | CH09LNL   |
| 42     | 376347  | RTA00002675F.1.08.1.P.Seq | F           | M00039249C:G11 | CH09LNL   |
| 43     | 446747  | RTA00002689F.d.16.2.P.Seq | F           | M00042740A:E09 | CH15CON   |
| 44     | 28092   | RTA00002711F.g.12.1.P.Seq | F           | M00023032A:B05 | CH03MAH   |
| 45     | 378206  | RTA00002671F.a.20.3.P.Seq | F           | M00033588C:G04 | CH09LNL   |
| 46     | 378206  | RTA00002671F.a.20.2.P.Seq | F           | M00033588C:G04 | CH09LNL   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 47     | 14940   | RTA00002709F.d.11.1.P.Seq | F           | M00005623A:G02 | CH02COH   |
| 48     | 378411  | RTA00002672F.g.13.2.P.Seq | F           | M00039004B:A06 | CH09LNL   |
| 49     | 38120   | RTA00002712F.i.14.1.P.Seq | F           | M00026927D:F02 | CH04MAL   |
| 50     | 375730  | RTA00002678F.l.13.2.P.Seq | F           | M00039612B:G05 | CH09LNL   |
| 51     | 428959  | RTA00002667F.h.15.1.P.Seq | F           | M00032811B:D02 | CH08LNH   |
| 52     | 376851  | RTA00002677F.c.03.2.P.Seq | F           | M00039341C:H11 | CH09LNL   |
| 53     | 373808  | RTA00002671F.d.14.2.P.Seq | F           | M00038272A:G01 | CH09LNL   |
| 54     | 376168  | RTA00002675F.n.17.1.P.Seq | F           | M00039258B:E06 | CH09LNL   |
| 55     | 18653   | RTA00002712F.o.08.1.P.Seq | F           | M00027135A:B11 | CH04MAL   |
| 56     | 187632  | RTA00002664F.i.15.1.P.Seq | F           | M00027617B:C12 | CH04MAL   |
| 57     | 374122  | RTA00002673F.l.22.1.P.Seq | F           | M00039104D:C09 | CH09LNL   |
| 58     | 374946  | RTA00002673F.j.24.1.P.Seq | F           | M00039096A:E07 | CH09LNL   |
| 59     | 375666  | RTA00002677F.n.16.2.P.Seq | F           | M00039422D:F04 | CH09LNL   |
| 60     | 162369  | RTA00002713F.d.24.1.P.Seq | F           | M00027292D:F10 | CH04MAL   |
| 61     | 21480   | RTA00002709F.c.18.2.P.Seq | F           | M00005531D:F06 | CH02COH   |
| 62     | 18560   | RTA00002711F.e.20.1.P.Seq | F           | M00022938B:F07 | CH03MAH   |
| 63     | 96575   | RTA00002663F.j.08.1.P.Seq | F           | M00022641C:H05 | CH03MAH   |
| 64     | 377576  | RTA00002682F.f.18.1.P.Seq | F           | M00039975C:C11 | CH09LNL   |
| 65     | 446747  | RTA00002689F.d.16.3.P.Seq | F           | M00042740A:E09 | CH15CON   |
| 66     | 379311  | RTA00002682F.g.01.1.P.Seq | F           | M00039976D:A12 | CH09LNL   |
| 67     | 379311  | RTA00002682F.f.24.1.P.Seq | F           | M00039976D:A12 | CH09LNL   |
| 68     | 124549  | RTA00002713F.c.07.1.P.Seq | F           | M00027237C:B08 | CH04MAL   |
| 69     | 449785  | RTA00002691F.c.07.3.P.Seq | F           | M00043345C:A06 | CH17COHLV |
| 70     | 375134  | RTA00002673F.k.22.2.P.Seq | F           | M00039099A:H08 | CH09LNL   |
| 71     | 186593  | RTA00002713F.n.15.1.P.Seq | F           | M00027620D:F11 | CH04MAL   |
| 72     | 449831  | RTA00002691F.a.17.3.P.Seq | F           | M00042518D:A06 | CH17COHLV |
| 73     | 379678  | RTA00002676F.b.06.1.P.Seq | F           | M00039274B:G07 | CH09LNL   |
| 74     | 20599   | RTA00002708F.h.06.1.P.Seq | F           | M00004264B:A05 | CH01COH   |
| 75     | 41115   | RTA00002713F.o.11.1.P.Seq | F           | M00027652B:F11 | CH04MAL   |
| 76     | 21109   | RTA00002708F.h.12.1.P.Seq | F           | M00004278A:F09 | CH01COH   |
| 77     | 455702  | RTA00002694F.b.11.1.P.Seq | F           | M00043433C:G07 | CH20COHLV |
| 78     | 380643  | RTA00002683F.p.09.2.P.Seq | F           | M00040103B:H10 | CH09LNL   |
| 79     | 374413  | RTA00002672F.i.15.2.P.Seq | F           | M00039015B:G10 | CH09LNL   |
| 80     | 378891  | RTA00002672F.i.18.2.P.Seq | F           | M00039016A:A02 | CH09LNL   |
| 81     | 379374  | RTA00002672F.k.11.2.P.Seq | F           | M00039028C:B11 | CH09LNL   |
| 82     | 17253   | RTA00002709F.h.23.1.P.Seq | F           | M00006866A:D07 | CH02COH   |
| 83     | 21565   | RTA00002709F.e.11.1.P.Seq | F           | M00005778B:F09 | CH02COH   |
| 84     | 373996  | RTA00002673F.n.11.1.P.Seq | F           | M00039108D:B06 | CH09LNL   |
| 85     | 380437  | RTA00002683F.c.09.1.P.Seq | F           | M00040039D:D06 | CH09LNL   |
| 86     | 430729  | RTA00002669F.h.18.2.P.Seq | F           | M00033226A:A11 | CH08LNH   |
| 87     | 376791  | RTA00002674F.l.17.1.P.Seq | F           | M00039166B:G06 | CH09LNL   |
| 88     | 373760  | RTA00002672F.p.20.1.P.Seq | F           | M00039049D:G07 | CH09LNL   |
| 89     | 373837  | RTA00002672F.p.22.1.P.Seq | F           | M00039050A:H10 | CH09LNL   |
| 90     | 376436  | RTA00002678F.h.17.2.P.Seq | F           | M00039476B:A02 | CH09LNL   |
| 91     | 373881  | RTA00002672F.b.20.1.P.Seq | F           | M00038638D:H03 | CH09LNL   |
| 92     | 377086  | RTA00002676F.p.07.1.P.Seq | F           | M00039328D:D07 | CH09LNL   |
| 93     | 377889  | RTA00002672F.c.08.1.P.Seq | F           | M00038661A:A07 | CH09LNL   |
| 94     | 380442  | RTA00002684F.b.05.2.P.Seq | F           | M00040111C:D05 | CH09LNL   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 95     | 374689  | RTA00002676F.m.13.2.P.Seq | F            | M00039318B:B09 | CH09LNL   |
| 96     | 375339  | RTA00002678F.m.23.2.P.Seq | F            | M00039616A:B10 | CH09LNL   |
| 97     | 14197   | RTA00002710F.f.15.1.P.Seq | F            | M00022084D:B01 | CH03MAH   |
| 98     | 380666  | RTA00002684F.c.04.2.P.Seq | F            | M00040115B:H12 | CH09LNL   |
| 99     | 377352  | RTA00002677F.i.13.2.P.Seq | F            | M00039404B:A05 | CH09LNL   |
| 100    | 379188  | RTA00002682F.a.03.1.P.Seq | F            | M00039914D:G12 | CH09LNL   |
| 101    | 428269  | RTA00002666F.c.13.1.P.Seq | F            | M00032539B:C11 | CH08LNH   |
| 102    | 373464  | RTA00002671F.l.13.3.P.Seq | F            | M00038327A:C11 | CH09LNL   |
| 103    | 15527   | RTA00002710F.p.07.1.P.Seq | F            | M00022747D:E03 | CH03MAH   |
| 104    | 377504  | RTA00002671F.i.17.3.P.Seq | F            | M00038303C:D02 | CH09LNL   |
| 105    | 33508   | RTA00002710F.g.17.1.P.Seq | F            | M00022183B:C02 | CH03MAH   |
| 106    | 129179  | RTA00002662F.d.19.2.P.Seq | F            | M00007157C:F11 | CH02COH   |
| 107    | 377086  | RTA00002676F.p.07.2.P.Seq | F            | M00039328D:D07 | CH09LNL   |
| 108    | 375872  | RTA00002675F.h.15.1.P.Seq | F            | M00039233A:A03 | CH09LNL   |
| 109    | 375652  | RTA00002676F.i.07.3.P.Seq | F            | M00039303C:F11 | CH09LNL   |
| 110    | 374266  | RTA00002674F.i.08.2.P.Seq | F            | M00039144C:E06 | CH09LNL   |
| 111    | 378983  | RTA00002682F.a.07.1.P.Seq | F            | M00039915D:C11 | CH09LNL   |
| 112    | 377343  | RTA00002684F.g.04.1.P.Seq | f            | M00040302C:A04 | CH09LNL   |
| 113    | 378679  | RTA00002681F.f.16.2.P.Seq | F            | M00039869B:F06 | CH09LNL   |
| 114    | 374095  | RTA00002671F.p.08.2.P.Seq | F            | M00038618C:C08 | CH09LNL   |
| 115    | 375843  | RTA00002671F.o.06.2.P.Seq | F            | M00038614C:H11 | CH09LNL   |
| 116    | 377788  | RTA00002684F.h.01.2.P.Seq | F            | M00040305C:H06 | CH09LNL   |
| 117    | 21403   | RTA00002709F.j.05.1.P.Seq | F            | M00006928D:D07 | CH02COH   |
| 118    | 23184   | RTA00002709F.b.05.2.P.Seq | F            | M00005358B:B06 | CH02COH   |
| 119    | 15671   | RTA00002710F.k.16.1.P.Seq | f            | M00022495D:H08 | CH03MAH   |
| 120    | 177367  | RTA00002663F.m.22.1.P.Seq | F            | M00022986D:H09 | CH03MAH   |
| 121    | 377788  | RTA00002684F.g.24.1.P.Seq | F            | M00040305C:H06 | CH09LNL   |
| 122    | 375058  | RTA00002675F.h.02.1.P.Seq | F            | M00039230D:G12 | CH09LNL   |
| 123    | 380412  | RTA00002680F.k.15.2.P.Seq | F            | M00039816B:D04 | CH09LNL   |
| 124    | 178447  | RTA00002663F.n.06.1.P.Seq | F            | M00023007A:H04 | CH03MAH   |
| 125    | 376647  | RTA00002674F.h.07.1.P.Seq | F            | M00039140D:D09 | CH09LNL   |
| 126    | 44679   | RTA00002661F.e.19.1.P.Seq | F            | M00003800A:F09 | CH01COH   |
| 127    | 377659  | RTA00002678F.a.04.2.P.Seq | F            | M00039430B:F12 | CH09LNL   |
| 128    | 379703  | RTA00002682F.h.03.1.P.Seq | F            | M00039982C:H04 | CH09LNL   |
| 129    | 374673  | RTA00002673F.e.08.2.P.Seq | F            | M00039068B:B04 | CH09LNL   |
| 130    | 20513   | RTA00002710F.j.12.1.P.Seq | F            | M00022391D:F10 | CH03MAH   |
| 131    | 376124  | RTA00002682F.n.17.1.P.Seq | F            | M00040021A:F09 | CH09LNL   |
| 132    | 374679  | RTA00002676F.d.07.2.P.Seq | F.           | M00039281D:B04 | CH09LNL   |
| 133    | 23184   | RTA00002709F.b.05.1.P.Seq | F            | M00005358B:B06 | CH02COH   |
| 134    | 430953  | RTA00002668F.i.23.1.P.Seq | F            | M00033007C:E01 | CH08LNH   |
| 135    | 380442  | RTA00002684F.b.05.1.P.Seq | F            | M00040111C:D05 | CH09LNL   |
| 136    | 12374   | RTA00002709F.a.01.1.P.Seq | F            | M00004825D:D05 | CH02COH   |
| 137    | 427466  | RTA00002665F.b.11.3.P.Seq | F            | M00028184D:G10 | CH08LNH   |
| 138    | 36611   | RTA00002668F.f.03.1.P.Seq | F            | M00032942D:C12 | CH08LNH   |
| 139    | 33756   | RTA00002662F.a.18.2.P.Seq | F            | M00005359A:D04 | CH02COH   |
| 140    | 456026  | RTA00002694F.e.03.1.P.Seq | F            | M00043616C:A05 | CH20COHLV |
| 141    | 15766   | RTA00002710F.k.02.1.P.Seq | F            | M00022444D:G01 | CH03MAH   |
| 142    | 24352   | RTA00002709F.a.05.1.P.Seq | F            | M00004839C:H02 | CH02COH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|--------------|----------------|---------|
| 143    | 24354   | RTA00002709F.a.03.1.P.Seq | F            | M00004832D:H02 | CH02COH |
| 144    | 379114  | RTA00002681F.o.01.2.P.Seq | F            | M00039903C:F03 | CH09LNL |
| 145    | 19609   | RTA00002709F.c.05.1.P.Seq | F            | M00005457C:A03 | CH02COH |
| 146    | 21685   | RTA00002709F.e.23.1.P.Seq | F            | M00006581D:F08 | CH02COH |
| 147    | 380085  | RTA00002682F.i.10.1.P.Seq | F            | M00039987A:F09 | CH09LNL |
| 148    | 20700   | RTA00002710F.i.18.1.P.Seq | F            | M00022373A:B05 | CH03MAH |
| 149    | 379981  | RTA00002682F.i.18.1.P.Seq | F            | M00039988A:E06 | CH09LNL |
| 150    | 376591  | RTA00002675F.c.01.1.P.Seq | F            | M00039213A:D01 | CH09LNL |
| 151    | 92058   | RTA00002663F.m.04.1.P.Seq | F            | M00022895A:H08 | CH03MAH |
| 152    | 196936  | RTA00002663F.m.02.1.P.Seq | F            | M00022885C:H05 | CH03MAH |
| 153    | 430702  | RTA00002668F.h.04.1.P.Seq | F            | M00032990B:A11 | CH08LNL |
| 154    | 378448  | RTA00002680F.n.21.2.P.Seq | F            | M00039832A:B12 | CH09LNL |
| 155    | 41606   | RTA00002713F.e.10.1.P.Seq | F            | M00027301A:G05 | CH04MAL |
| 156    | 213817  | RTA00002665F.i.19.1.P.Seq | F            | M00027634A:D11 | CH04MAL |
| 157    | 373464  | RTA00002671F.l.13.1.P.Seq | F            | M00038327A:C11 | CH09LNL |
| 158    | 379483  | RTA00002679F.k.12.1.P.Seq | F            | M00039700B:D02 | CH09LNL |
| 159    | 375796  | RTA00002680F.f.17.1.P.Seq | F            | M00039795B:H10 | CH09LNL |
| 160    | 375796  | RTA00002680F.f.17.2.P.Seq | F            | M00039795B:H10 | CH09LNL |
| 161    | 170485  | RTA00002663F.b.12.1.P.Seq | F            | M00021665B:F12 | CH03MAH |
| 162    | 374291  | RTA00002673F.f.17.1.P.Seq | F            | M00039072C:E02 | CH09LNL |
| 163    | 380513  | RTA00002677F.p.15.2.P.Seq | F            | M00039428C:E01 | CH09LNL |
| 164    | 379416  | RTA00002683F.j.07.2.P.Seq | F            | M00040077D:C11 | CH09LNL |
| 165    | 378178  | RTA00002680F.l.13.1.P.Seq | F            | M00039820A:F11 | CH09LNL |
| 166    | 427947  | RTA00002665F.n.24.1.P.Seq | F            | M00032495B:D02 | CH08LNL |
| 167    | 427269  | RTA00002665F.d.03.3.P.Seq | F            | M00028212C:B08 | CH08LNL |
| 168    | 20451   | RTA00002710F.j.10.1.P.Seq | F            | M00022391B:E01 | CH03MAH |
| 169    | 377003  | RTA00002683F.g.09.2.P.Seq | F            | M00040062B:B05 | CH09LNL |
| 170    | 427759  | RTA00002665F.o.11.1.P.Seq | F            | M00032499C:A01 | CH08LNL |
| 171    | 427549  | RTA00002668F.k.13.1.P.Seq | F            | M00033034C:A06 | CH08LNL |
| 172    | 373881  | RTA00002672F.b.20.2.P.Seq | F            | M00038638D:H03 | CH09LNL |
| 173    | 188215  | RTA00002664F.f.13.2.P.Seq | F            | M00027200A:F02 | CH04MAL |
| 174    | 397683  | RTA00002681F.d.04.2.P.Seq | F            | M00039857B:G10 | CH09LNL |
| 175    | 380652  | RTA00002679F.d.12.2.P.Seq | F            | M00039455D:H04 | CH09LNL |
| 176    | 378334  | RTA00002678F.h.10.1.P.Seq | F            | M00039682C:H11 | CH09LNL |
| 177    | 377930  | RTA00002680F.g.14.1.P.Seq | F            | M00039798B:B02 | CH09LNL |
| 178    | 378692  | RTA00002680F.o.20.3.P.Seq | F            | M00039835A:F07 | CH09LNL |
| 179    | 32279   | RTA00002709F.d.23.1.P.Seq | F            | M00005673B:B12 | CH02COH |
| 180    | 376379  | RTA00002680F.c.15.1.P.Seq | F            | M00039782A:H10 | CH09LNL |
| 181    | 375963  | RTA00002675F.i.12.1.P.Seq | F            | M00039238A:B12 | CH09LNL |
| 182    | 378683  | RTA00002680F.a.14.2.P.Seq | F            | M00039773D:A09 | CH09LNL |
| 183    | 374946  | RTA00002673F.j.24.2.P.Seq | F            | M00039096A:E07 | CH09LNL |
| 184    | 429583  | RTA00002666F.g.10.1.P.Seq | F            | M00032584A:H08 | CH08LNL |
| 185    | 28338   | RTA00002711F.e.17.1.P.Seq | F            | M00029930C:E02 | CH03MAH |
| 186    | 427970  | RTA00002665F.j.13.1.P.Seq | F            | M00031368A:E10 | CH08LNL |
| 187    | 379650  | RTA00002683F.h.22.2.P.Seq | F            | M00040072C:G09 | CH09LNL |
| 188    | 379661  | RTA00002676F.c.05.2.P.Seq | F            | M00039277D:G10 | CH09LNL |
| 189    | 376182  | RTA00002677F.b.17.2.P.Seq | F            | M00039340B:E07 | CH09LNL |
| 190    | 374797  | RTA00002678F.b.12.2.P.Seq | F            | M00039444C:H02 | CH09LNL |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 191    | 375389  | RTA00002674F.a.13.1.P.Seq | F           | M00038120C:C09 | CH09LNL   |
| 192    | 397115  | RTA00002683F.i.22.2.P.Seq | F           | M00040076C:D06 | CH09LNL   |
| 193    | 186655  | RTA00002712F.i.21.1.P.Seq | F           | M00026941D:A04 | CH04MAL   |
| 194    | 404682  | RTA00002687F.b.13.1.P.Seq | F           | M00039766D:H01 | CH14EDT   |
| 195    | 19609   | RTA00002709F.c.05.2.P.Seq | F           | M00005457C:A03 | CH02COH   |
| 196    | 404682  | RTA00002687F.b.13.2.P.Seq | F           | M00039766D:H01 | CH14EDT   |
| 197    | 380412  | RTA00002680F.k.15.1.P.Seq | F           | M00039816B:D04 | CH09LNL   |
| 198    | 394413  | RTA00002689F.d.17.3.P.Seq | F           | M00042742D:D05 | CH15CON   |
| 199    | 380086  | RTA00002679F.m.16.1.P.Seq | F           | M00039710C:G03 | CH09LNL   |
| 200    | 430738  | RTA00002669F.i.1652.P.Seq | F           | M00033231D:B09 | CH08LNH   |
| 201    | 40667   | RTA00002712F.g.22.1.P.Seq | F           | M00026882D:G09 | CH04MAL   |
| 202    | 397421  | RTA00002681F.c.16.2.P.Seq | F           | M00039854B:F09 | CH09LNL   |
| 203    | 398775  | RTA00002679F.f.11.1.P.Seq | F           | M00039675D:H05 | CH09LNL   |
| 204    | 87345   | RTA00002712F.f.19.1.P.Seq | F           | M00026850D:F09 | CH04MAL   |
| 205    | 379920  | RTA00002679F.b.24.2.P.Seq | F           | M00039660C:C10 | CH09LNL   |
| 206    | 380666  | RTA00002684F.c.04.1.P.Seq | F           | M00040115B:H12 | CH09LNL   |
| 207    | 404340  | RTA00002687F.b.05.2.P.Seq | F           | M00039764C:D07 | CH14EDT   |
| 208    | 375509  | RTA00002860F.e.08.2.P.Seq | F           | M00039790B:D03 | CH09LNL   |
| 209    | 46423   | RTA00002712F.i.02.1.P.Seq | F           | M00026914A:H10 | CH04MAL   |
| 210    | 401713  | RTA00002685F.p.10.2.P.Seq | F           | M00039647A:H11 | CH12EDT   |
| 211    | 377003  | RTA00002683F.g.09.1.P.Seq | F           | M00040062B:B05 | CH09LNL   |
| 212    | 378891  | RTA00002672F.i.18.1.P.Seq | F           | M00039016A:A02 | CH09LNL   |
| 213    | 412778  | RTA00002685F.i.07.2.P.Seq | F           | M00039533D:F04 | CH12EDT   |
| 214    | 373786  | RTA00002679F.a.20.2.P.Seq | F           | M00039655C:C07 | CH09LNL   |
| 215    | 378692  | RTA00002680F.o.20.2.P.Seq | F           | M00039835A:F07 | CH09LNL   |
| 216    | 88888   | RTA00002713F.f.22.1.P.Seq | F           | M00027355A:B07 | CH04MAL   |
| 217    | 358187  | RTA00002676F.b.04.2.P.Seq | F           | M00039273D:B02 | CH09LNL   |
| 218    | 377131  | RTA00002682F.e.10.1.P.Seq | F           | M00039938C:E11 | CH09LNL   |
| 219    | 21488   | RTA00002708F.f.17.1.P.Seq | F           | M00004152A:C12 | CH01COH   |
| 220    | 447487  | RTA00002689F.e.04.3.P.Seq | F           | M00042895A:D10 | CH15CON   |
| 221    | 364     | RTA00002710F.a.06.1.P.Seq | F           | M00007929C:B08 | CH03MAH   |
| 222    | 404024  | RTA00002687F.e.18.2.P.Seq | F           | M00039958A:A08 | CH14EDT   |
| 223    | 152305  | RTA00002712F.d.02.1.P.Seq | F           | M00023376B:G04 | CH04MAL   |
| 224    | 106050  | RTA00002713F.o.17.1.P.Seq | F           | M00027668C:H12 | CH04MAL   |
| 225    | 41126   | RTA00002713F.l.12.1.P.Seq | F           | M00027546C:B10 | CH04MAL   |
| 226    | 113496  | RTA00002713F.n.20.1.P.Seq | F           | M00027625A:H01 | CH04MAL   |
| 227    | 447487  | RTA00002689F.e.04.1.P.Seq | F           | M00042895A:D10 | CH15CON   |
| 228    | 146335  | RTA00002712F.j.17.1.P.Seq | F           | M00026980A:D09 | CH04MAL   |
| 229    | 376647  | RTA00002674F.h.07.2.P.Seq | F           | M00039140D:D09 | CH09LNL   |
| 230    | 376746  | RTA00002674F.f.12.2.P.Seq | F           | M00039133B:F08 | CH09LNL   |
| 231    | 373523  | RTA00002674F.n.21.2.P.Seq | F           | M00039177B:D03 | CH09LNL   |
| 232    | 455466  | RTA00002694F.c.10.1.P.Seq | F           | M00043461D:E06 | CH20COHLV |
| 233    | 374031  | RTA00002683F.p.17.2.P.Seq | F           | M00040105C:F14 | CH09LNL   |
| 234    | 373997  | RTA00002673F.m.04.2.P.Seq | F           | M00039105C:B08 | CH09LNL   |
| 235    | 455717  | RTA00002694F.a.06.1.P.Seq | F           | M00042593C:G06 | CH20COHLV |
| 236    | 373837  | RTA00002672F.p.22.2.P.Seq | F           | M00039050A:H10 | CH09LNL   |
| 237    | 374513  | RTA00002672F.i.16.2.P.Seq | F           | M00039015B:H09 | CH09LNL   |
| 238    | 375628  | RTA00002672F.k.04.2.P.Seq | F           | M00039026D:F05 | CH09LNL   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|--------------|----------------|---------|
| 239    | 377732  | RTA00002681F.p.09.1.P.Seq | F            | M00039910C:G10 | CH09LNL |
| 240    | 378326  | RTA00002681F.m.11.1.P.Seq | F            | M00039896C:H01 | CH09LNL |
| 241    | 378001  | RTA00002681F.m.22.1.P.Seq | F            | M00039898D:C06 | CH09LNL |
| 242    | 378459  | RTA00002681F.i.07.2.P.Seq | F            | M00039879D:B11 | CH09LNL |
| 243    | 373862  | RTA00002671F.g.01.2.P.Seq | F            | M00038284B:H04 | CH09LNL |
| 244    | 373252  | RTA00002670F.k.16.1.P.Seq | F            | M00033451A:H01 | CH09LNL |
| 245    | 378475  | RTA00002672F.g.24.1.P.Seq | F            | M00039006D:B01 | CH09LNL |
| 246    | 379941  | RTA00002682F.j.15.1.P.Seq | F            | M00039990C:D10 | CH09LNL |
| 247    | 427703  | RTA00002665F.e.11.1.P.Seq | F            | M00028357A:G10 | CH08LNH |
| 248    | 373976  | RTA00002671F.p.15.2.P.Seq | F            | M00038619B:A03 | CH09LNL |
| 249    | 431643  | RTA00002669F.l.16.3.P.Seq | F            | M00033276D:H09 | CH08LNH |
| 250    | 383502  | RTA00002670F.k.07.1.P.Seq | F            | M00033446D:B02 | CH09LNL |
| 251    | 378764  | RTA00002681F.j.04.1.P.Seq | F            | M00039884A:H11 | CH09LNL |
| 252    | 431629  | RTA00002669F.l.14.3.P.Seq | F            | M00033276B:G08 | CH08LNH |
| 253    | 372992  | RTA00002671F.b.16.2.P.Seq | F            | M00033594C:B03 | CH09LNL |
| 254    | 431601  | RTA00002669F.k.08.3.P.Seq | F            | M00033263B"G04 | CH08LHN |
| 255    | 21059   | RTA00002710F.c.05.1.P.Seq | F            | M00008053A:F10 | CH03MAH |
| 256    | 430689  | RTA00002669F.i.24.3.P.Seq | F            | M00033243B:A05 | CH08LNH |
| 257    | 131764  | RTA00002662F.c.14.1.P.Seq | F            | M00006893C:E07 | CH02COH |
| 258    | 373300  | RTA00002674F.c.21.2.P.Seq | F            | M00039126D:A08 | CH09LNL |
| 259    | 384601  | RTA00002670F.k.06.1.P.Seq | F            | M00033446C:G08 | CH09LNL |
| 260    | 375389  | RTA00002674F.a.13.2.P.Seq | F            | M00039120C:C09 | CH09LNL |
| 261    | 15248   | RTA00002710F.f.23.1.P.Seq | F            | M00022127C:H03 | CH03MAH |
| 262    | 428134  | RTA00002666F.c.15.1.P.Seq | F            | M00032540A:A09 | CH08LNH |
| 263    | 374184  | RTA00002672F.a.19.1.P.Seq | F            | M00038633A:D07 | CH09LNL |
| 264    | 136225  | RTA00002676F.n.02.2.P.Seq | F            | M00039319C:A04 | CH09LNL |
| 265    | 401713  | RTA00002685F.p.10.1.P.Seq | F            | M00039647A:H11 | CH12EDT |
| 266    | 27104   | RTA00002661F.a.09.1.P.Seq | F            | M00001363D:D09 | CH01COH |
| 267    | 207466  | RTA00002664F.j.08.2.P.Seq | F            | M00027733A:A02 | CH04MAL |
| 268    | 143045  | RTA00002663F.a.02.1.P.Seq | F            | M00007941D:C09 | CH03MAH |
| 269    | 378830  | RTA00002675F.e.07.1.P.Seq | F            | M00039221A:H03 | CH09LNL |
| 270    | 21731   | RTA00002709F.k.07.1.P.Seq | F            | M00007013A:D09 | CH02COH |
| 271    | 428552  | RTA00002666F.c.16.1.P.Seq | F            | M00032541D:H08 | CH08LNH |
| 272    | 187632  | RTA00002664F.i.15.2.P.Seq | F            | M00027617B:C12 | CH04MAL |
| 273    | 431053  | RTA00002668F.o.05.2.P.Seq | F            | M00033130B:F06 | CH08LNH |
| 274    | 188972  | RTA00002664F.d.20.1.P.Seq | F            | M00027030C:H06 | CH04MAL |
| 275    | 430678  | RTA00002668F.h.12.1.P.Seq | F            | M00032994A:A08 | CH08LNH |
| 276    | 374042  | RTA00002672F.a.08.1.P.Seq | F            | M00038631C:B10 | CH09LNL |
| 277    | 24332   | RTA00002709F.j.07.1.P.Seq | F            | M00006935C:F06 | CH02COH |
| 278    | 376764  | RTA00002674F.f.20.1.P.Seq | F            | M00039135D:F05 | CH09LNL |
| 279    | 134338  | RTA00002662F.c.15.2.P.Seq | F            | M00006897A:H02 | CH02COH |
| 280    | 375541  | RTA00002680F.d.21.2.P.Seq | F            | M00039788A:E03 | CH09LNL |
| 281    | 228909  | RTA00002664F.e.08.2.P.Seq | F            | M00027085C:E11 | CH04MAL |
| 282    | 58063   | RTA00002661F.h.18.1.P.Seq | F            | M00004234A:E07 | CH01COH |
| 283    | 380500  | RTA00002670F.p.19.2.P.Seq | F            | M00035583B:E06 | CH09LNL |
| 284    | 34928   | RTA00002710F.p.21.1.P.Seq | F            | M00022795B:G06 | CH03MAH |
| 285    | 374028  | RTA00002674F.k.03.2.P.Seq | F            | M00039156A:B11 | CH09LNL |
| 286    | 374121  | RTA00002672F.h.22.2.P.Seq | F            | M00039013A:C09 | CH09LNL |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|--------------|----------------|---------|
| 287    | 429547  | RTA00002668F.c.07.1.P.Seq | F            | M00032917D:G09 | CH08LNH |
| 288    | 380668  | RTA00002670F.p.11.2.P.Seq | F            | M00033581C:H10 | CH09LNL |
| 289    | 258704  | RTA00002665F.m.06.1.P.Seq | F            | M00032480B:E10 | CH08LNH |
| 290    | 380325  | RTA00002670F.p.22.2.P.Seq | F            | M00033583D:B05 | CH09LNL |
| 291    | 378326  | RTA00002681F.m.11.2.P.Seq | F            | M00039896C:H01 | CH09LNL |
| 292    | 375618  | RTA00002675F.d.13.1.P.Seq | F            | M00039218A:F03 | CH09LNL |
| 293    | 20999   | RTA00002709F.j.16.1.P.Seq | F            | M00006977C:G04 | CH02COH |
| 294    | 29102   | RTA00002710F.p.18.1.P.Seq | F            | M00022793D:B01 | CH03MAH |
| 295    | 379334  | RTA00002680F.b.22.1.P.Seq | F            | M00039778C:A04 | CH09LNL |
| 296    | 23943   | RTA00002709F.i.12.1.P.Seq | F            | M00006886D:H02 | CH02COH |
| 297    | 373998  | RTA00002672F.a.10.2.P.Seq | F            | M00038631D:B02 | CH09LNL |
| 298    | 373325  | RTA00002672F.c.14.2.P.Seq | F            | M00038662B:A12 | CH09LNL |
| 299    | 373818  | RTA00002672F.e.15.2.P.Seq | F            | M00038995C:G08 | CH09LNL |
| 300    | 429843  | RTA00002668F.c.10.1.P.Seq | F            | M00032918B:E06 | CH08LNH |
| 301    | 427755  | RTA00002665F.d.19.3.P.Seq | F            | M00028316B:H12 | CH08LNH |
| 302    | 189177  | RTA00002664F.c.23.2.P.Seq | F            | M00026922C:G03 | CH04MAL |
| 303    | 13294   | RTA00002709F.j.15.1.P.Seq | F            | M00006968A:G08 | CH02COH |
| 304    | 178801  | RTA00002663F.n.01.1.P.Seq | F            | M00022997A:F06 | CH03MAH |
| 305    | 230865  | RTA00002664F.d.03.2.P.Seq | F            | M00026928D:A03 | CH04MAL |
| 306    | 178801  | RTA00002663F.m.24.1.P.Seq | F            | M00022997A:F06 | CH03MAH |
| 307    | 378808  | RTA00002672F.g.21.2.P.Seq | F            | M00039005C:H01 | CH09LNL |
| 308    | 378957  | RTA00002670F.d.17.2.P.Seq | F            | M00033362C:C05 | CH09LNL |
| 309    | 373523  | RTA00002674F.n.21.1.P.Seq | F            | M00039177B:D03 | CH09LNL |
| 310    | 375458  | RTA00002678F.l.06.2.P.Seq | F            | M00039611D:D11 | CH09LNL |
| 311    | 429794  | RTA00002668F.c.09.1.P.Seq | F            | M00032918B:D08 | CH08LNH |
| 312    | 72797   | RTA00002661F.e.07.1.P.Seq | F            | M00003761C:F02 | CH01COH |
| 313    | 429992  | RTA00002668F.c.21.1.P.Seq | F            | M00032921B:H08 | CH08LNH |
| 314    | 374410  | RTA00002674F.k.11.2.P.Seq | F            | M00039158B:G12 | CH09LNL |
| 315    | 376553  | RTA00002674F.g.19.1.P.Seq | F            | M00039139A:C09 | CH09LNL |
| 316    | 429096  | RTA00002666F.f.16.1.P.Seq | F            | M00032578A:G06 | CH08LNH |
| 317    | 181948  | RTA00002663F.n.05.1.P.Seq | F            | M00023003C:D07 | CH03MAH |
| 318    | 378475  | RTA00002672F.h.01.2.P.Seq | F            | M00039006D:B01 | CH09LNL |
| 319    | 427336  | RTA00002665F.c.23.1.P.Seq | F            | M00028210B:D02 | CH08LNH |
| 320    | 374042  | RTA00002672F.a.08.2.P.Seq | F            | M00038631C:B10 | CH09LNL |
| 321    | 386543  | RTA00002672F.f.13.2.P.Seq | F            | M00038999B:G11 | CH09LNL |
| 322    | 376659  | RTA00002678F.h.11.2.P.Seq | F            | M00039475C:E10 | CH09LNL |
| 323    | 29135   | RTA00002663F.c.09.1.P.Seq | F            | M00021923C:D11 | CH03MAH |
| 324    | 377967  | RTA00002681F.m.17.2.P.Seq | F            | M00039897D:C10 | CH09LNL |
| 325    | 431330  | RTA00002668F.m.16.2.P.Seq | F            | M00033074A:C08 | CH08LNH |
| 326    | 373824  | RTA00002680F.i.19.2.P.Seq | F            | M00039808D:H02 | CH09LNL |
| 327    | 50094   | RTA00002661F.j.02.2.P.Seq | F            | M00004378A:B10 | CH01COH |
| 328    | 214272  | RTA00002664F.h.03.2.P.Seq | F            | M00027366A:F11 | CH04MAL |
| 329    | 377293  | RTA00002680F.b.17.2.P.Seq | F            | M00039777C:E05 | CH09LNL |
| 330    | 195053  | RTA00002663F.n.16.1.P.Seq | F            | M00023044B:D02 | CH03MAH |
| 331    | 21274   | RTA00002709F.m.09.1.P.Seq | F            | M00007194A:B09 | CH02COH |
| 332    | 376580  | RTA00002675F.b.20.1.P.Seq | F            | M00039212C:C12 | CH09LNL |
| 333    | 374725  | RTA00002673F.f.02.2.P.Seq | F            | M00039070D:C02 | CH09LNL |
| 334    | 25238   | RTA00002710F.n.08.1.P.Seq | F            | M00022634D:C08 | CH03MAH |



| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|-------------|-----------------|---------|
| 335    | 377337  | RTA00002683F.l.07.2.P.Seq | F           | M00040085D:A10  | CH09LNL |
| 336    | 450485  | RTA00002692F.a.13.2.P.Seq | F           | M00042625C:B04  | CH18CON |
| 337    | 21989   | RTA00002709F.h.22.1.P.Seq | F           | M00006861B:F09  | CH02COH |
| 338    | 58296   | RTA00002661F.i.20.2.P.Seq | F           | M00004354D:E05  | CH01COH |
| 339    | 379144  | RTA00002679F.l.14.1.P.Seq | F           | M00039705D:F02  | CH09LNL |
| 340    | 379690  | RTA00002680F.b.21.2.P.Seq | F           | M00039778B:G03  | CH09LNL |
| 341    | 379640  | RTA00002681F.d.12.2.P.Seq | F           | M00039859C:G10  | CH09LNL |
| 342    | 373988  | RTA00002673F.h.23.1.P.Seq | F           | M00039079A:A05  | CH09LNL |
| 343    | 373988  | RTA00002673F.h.23.2.P.Seq | F           | M00039079A:A05  | CH09LNL |
| 344    | 380673  | RTA00002673F.j.13.2.P.Seq | F           | M00039084C:H03  | CH09LNL |
| 345    | 55243   | RTA00002661F.i.06.2.P.Seq | F           | M00004282D:C11  | CH01COH |
| 346    | 40557   | RTA00002713F.h.21.1.P.Seq | F           | M00027398C:F07  | CH04MAL |
| 347    | 375467  | RTA00002677F.m.03.1.P.Seq | F           | M000039417A:D03 | CH09LNL |
| 348    | 398406  | RTA00002679F.j.02.1.P.Seq | F           | M00039689C:E08  | CH09LNL |
| 349    | 430392  | RTA00002668F.k.19.1.P.Seq | F           | M00033037D:C11  | CH08LNH |
| 350    | 376746  | RTA00002674F.f.12.1.P.Seq | F           | M00039133B:F08  | CH09LNL |
| 351    | 115595  | RTA00002713F.e.07.1.P.Seq | F           | M00027297A:C04  | CH04MAL |
| 352    | 377182  | RTA00002682F.l.11.1.P.Seq | F           | M00040010A:F10  | CH09LNL |
| 353    | 380659  | RTA00002684F.e.07.2.P.Seq | F           | M00040124D:H01  | CH09LNL |
| 354    | 373862  | RTA00002671F.g.01.1.P.Seq | F           | M00038284B:H04  | CH09LNL |
| 355    | 376096  | RTA00002677F.b.16.2.P.Seq | F           | M00039340A:D05  | CH09LNL |
| 356    | 372887  | RTA00002670F.d.05.2.P.Seq | F           | M00033358A:H12  | CH09LNL |
| 357    | 378475  | RTA00002672F.g.24.2.P.Seq | F           | M00039006D:B01  | CH09LNL |
| 358    | 427336  | RTA00002665F.c.23.3.P.Seq | F           | M00028210B:D02  | CH08LNH |
| 359    | 373814  | RTA00002672F.b.02.2.P.Seq | F           | M00038635A:G09  | CH09LNL |
| 360    | 215506  | RTA00002664F.h.08.2.P.Seq | F           | M00027438C:G07  | CH04MAL |
| 361    | 374465  | RTA00002673F.c.07.2.P.Seq | F           | M00039058C:H02  | CH09LNL |
| 362    | 428784  | RTA00002667F.c.18.1.P.Seq | F           | M00032744B:F10  | CH08LNH |
| 363    | 379581  | RTA00002676F.a.21.2.P.Seq | F           | M00039273B:F02  | CH09LNL |
| 364    | 378371  | RTA00002678F.f.20.2.P.Seq | F           | M00039465A:A08  | CH09LNL |
| 365    | 375154  | RTA00002676F.c.13.2.P.Seq | F           | M00039279B:H02  | CH09LNL |
| 366    | 431214  | RTA00002669F.k.04.1.P.Seq | F           | M00033262D:A11  | CH08LNH |
| 367    | 376053  | RTA00002675F.l.03.1.P.Seq | F           | M00039249A:C12  | CH09LNL |
| 368    | 373282  | RTA00002680F.j.19.2.P.Seq | F           | M00039813B:D11  | CH09LNL |
| 369    | 33397   | RTA00002661F.h.04.1.P.Seq | F           | M00004168A:G11  | CH01COH |
| 370    | 376706  | RTA00002675F.c.02.1.P.Seq | F           | M00039213B:F05  | CH09LNL |
| 371    | 378292  | RTA00002681F.i.09.2.P.Seq | F           | M00039880A:H11  | CH09LNL |
| 372    | 431612  | RTA00002669F.e.23.3.P.Seq | F           | M00033202D:G06  | CH08LNH |
| 373    | 378471  | RTA00002679F.o.17.1.P.Seq | F           | M00039727C:B09  | CH09LNL |
| 374    | 378666  | RTA00002681F.i.05.2.P.Seq | F           | M00039879C:F05  | CH09LNL |
| 375    | 374894  | RTA00002675F.f.04.1.P.Seq | F           | M00039224A:E12  | CH09LNL |
| 376    | 430191  | RTA00002667F.j.24.1.P.Seq | F           | M00032829B:E06  | CH08LNH |
| 377    | 428581  | RTA00002667F.c.12.1.P.Seq | F           | M00032739A:A06  | CH08LNH |
| 378    | 379598  | RTA00002679F.k.03.1.P.Seq | F           | M00039697B:F11  | CH09LNL |
| 379    | 45300   | RTA00002710F.j.23.1.P.Seq | F           | M00022434D:D06  | CH03MAH |
| 380    | 23030   | RTA00002709F.b.10.1.P.Seq | F           | M00005384A:C11  | CH02COH |
| 381    | 379928  | RTA00002679F.o.06.1.P.Seq | F           | M00032829B:E06  | CH08LNH |
| 382    | 430191  | RTA00002667F.k.01.1.P.Seq | F           | M00039720D:D02  | CH09LNL |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|--------------|----------------|---------|
| 383    | 374685  | RTA00002675F.g.02.1.P.Seq | F            | M00039228A:B05 | CH09LNL |
| 384    | 375728  | RTA00002676F.h.05.2.P.Seq | F            | M00039299B:G12 | CH09LNL |
| 385    | 230237  | RTA00002670F.b.08.2.P.Seq | F            | M00033306D:H09 | CH09LNL |
| 386    | 380673  | RTA00002673F.j.13.1.P.Seq | F            | M00039084C:H03 | CH09LNL |
| 387    | 378938  | RTA00002679F.k.20.1.P.Seq | F            | M00039702A:B12 | CH09LNL |
| 388    | 375115  | RTA00002673F.e.01.1.P.Seq | F            | M00039066D:G08 | CH09LNL |
| 389    | 378673  | RTA00002680F.p.21.2.P.Seq | F            | M00039838A:F05 | CH09LNL |
| 390    | 372909  | RTA00002670F.a.12.2.P.Seq | F            | M00033300D:H12 | CH09LNL |
| 391    | 373300  | RTA00002674F.c.21.1.P.Seq | F            | M00039126D:A08 | CH09LNL |
| 392    | 379318  | RTA00002683F.h.16.2.P.Seq | F            | M00040071B:A10 | CH09LNL |
| 393    | 378319  | RTA00002681F.k.07.2.P.Seq | F            | M00039890A:H05 | CH09LNL |
| 394    | 374608  | RTA00002675F.g.20.1.P.Seq | F            | M00039230A:A10 | CH09LNL |
| 395    | 374328  | RTA00002673F.c.24.2.P.Seq | F            | M00039061B:F08 | CH09LNL |
| 396    | 374328  | RTA00002673F.d.01.2.P.Seq | F            | M00039061B:F08 | CH09LNL |
| 397    | 428401  | RTA00002667F.b.07.1.P.Seq | F            | M00032725C:F06 | CH08LNH |
| 398    | 136202  | RTA00002687F.p.05.2.P.Seq | F            | M00040349D:B09 | CH14EDT |
| 399    | 374394  | RTA00002673F.c.15.1.P.Seq | F            | M00039059C:G08 | CH09LNL |
| 400    | 37784   | RTA00002708F.c.17.1.P.Seq | F            | M00003816D:E11 | CH01COH |
| 401    | 378282  | RTA00002681F.h.11.1.P.Seq | F            | M00039876D:H09 | CH09LNL |
| 402    | 185663  | RTA00002712F.p.17.2.P.Seq | F            | M00027178B:G09 | CH04MAL |
| 403    | 14866   | RTA00002709F.d.14.1.P.Seq | F            | M00005623D:G12 | CH02COH |
| 404    | 383502  | RTA00002670F.k.07.2.P.Seq | F            | M00033446D:B02 | CH09LNL |
| 405    | 13463   | RTA00002709F.f.18.1.P.Seq | F            | M00006657C:G05 | CH02COH |
| 406    | 21274   | RTA00002709F.m.09.2.P.Seq | F            | M00007194A:B09 | CH02COH |
| 407    | 13745   | RTA00002714F.b.13.1.P.Seq | F            | M00027801C:C11 | CH04MAL |
| 408    | 23485   | RTA00002714F.c.10.1.P.Seq | F            | M00027836D:F12 | CH04MAL |
| 409    | 428364  | RTA00002667F.c.09.1.P.Seq | F            | M00032737B:E09 | CH08LNH |
| 410    | 431629  | RTA00002669F.l.14.2.P.Seq | F            | M00033276B:G08 | CH08LNH |
| 411    | 379754  | RTA00002682F.h.08.1.P.Seq | F            | M00039983D:A06 | CH09LNL |
| 412    | 431601  | RTA00002669F.k.08.2.P.Seq | F            | M00033263B:G04 | CH08LNH |
| 413    | 375749  | RTA00002680F.f.23.2.P.Seq | F            | M00039795D:G06 | CH09LNL |
| 414    | 378764  | RTA00002681F.j.04.2.P.Seq | F.           | M00039884A:H11 | CH09LNL |
| 415    | 215605  | RTA00002664F.i.20.1.P.Seq | F            | M00027647C:D03 | CH04MAL |
| 416    | 376144  | RTA00002675F.j.09.1.P.Seq | F            | M0003941A:E11  | CH09LNL |
| 417    | 373071  | RTA00002670F.j.23.2.P.Seq | F            | M00033442A:D06 | CH09LNL |
| 418    | 379684  | RTA00002681F.c.09.2.P.Seq | F            | M00039851B:G11 | CH09LNL |
| 419    | 379610  | RTA00002680F.k.11.2.P.Seq | F            | M00039815C:F09 | CH09LNL |
| 420    | 22392   | RTA00002708F.a.10.1.P.Seq | F            | M00001395D:H02 | CH01COH |
| 421    | 377555  | RTA00002683F.l.08.2.P.Seq | F            | M00040085D:E04 | CH09LNL |
| 422    | 32624   | RTA00002713F.f.15.1.P.Seq | F            | M00027347C:G07 | CH04MAL |
| 423    | 375024  | RTA00002675F.p.12.1.P.Seq | F            | M00039266D:F12 | CH09LNL |
| 424    | 374725  | RTA00002673F.f.02.1.P.Seq | F            | M00039070D:C02 | CH09LNL |
| 425    | 376228  | RTA00002676F.f.19.2.P.Seq | F            | M00039293A:H04 | CH09LNL |
| 426    | 375906  | RTA00002675F.i.18.1.P.Seq | F            | M00039238D:A08 | CH09LNL |
| 427    | 186190  | RTA00002714F.a.04.1.P.Seq | F            | M00027729D:H06 | CH04MAL |
| 428    | 57694   | RTA00002713F.f.02.1.P.Seq | F            | M00027319D:B11 | CH04MAL |
| 429    | 7007    | RTA00002709F.d.08.1.P.Seq | F            | M00005614B:B01 | CH02COH |
| 430    | 40084   | RTA00002685F.o.19.2.P.Seq | F            | M00039641C:D07 | CH12EDT |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 431    | 375648  | RTA00002676F.h.18.2.P.Seq | F           | M00039301B:F06 | CH09LNL   |
| 432    | 166493  | RTA00002663F.h.08.1.P.Seq | F           | M00022492C:A02 | CH03MAH   |
| 433    | 379632  | RTA00002682F.h.14.1.P.Seq | F           | M00039984B:G12 | CH09LNL   |
| 434    | 373234  | RTA00002676F.g.15.2.P.Seq | F           | M00039297C:H08 | CH09LNL   |
| 435    | 401230  | RTA00002685F.i.05.2.P.Seq | F           | M00039533A:C12 | CH12EDT   |
| 436    | 186623  | RTA00002712F.f.15.1.P.Seq | F           | M00026843B:D10 | CH04MAL   |
| 437    | 127714  | RTA00002712F.k.14.1.P.Seq | F           | M00027018A:C09 | CH04MAL   |
| 438    | 451857  | RTA00002692F.a.01.1.P.Seq | F           | M00042584B:C10 | CH18CON   |
| 439    | 404620  | RTA00002687F.c.03.2.P.Seq | F           | M00039770A:G11 | CH14EDT   |
| 440    | 186872  | RTA00002663F.k.23.1.P.Seq | F           | M00022797B:G08 | CH03MAH   |
| 441    | 42729   | RTA00002709F.c.06.2.P.Seq | F           | M00005458A:F11 | CH02COH   |
| 442    | 373380  | RTA00002674F.b.07.1.P.Seq | F           | M00039123A:B10 | CH09LNL   |
| 443    | 374465  | RTA00002673F.c.07.1.P.Seq | F           | M00039058C:H02 | CH09LNL   |
| 444    | 403557  | RTA00002687F.d.10.2.P.Seq | F           | M00039948A:E03 | CH14EDT   |
| 445    | 16749   | RTA00002709F.b.14.2.P.Seq | F           | M00005402B:F08 | CH02COH   |
| 446    | 375592  | RTA00002680F.f.22.2.P.Seq | F           | M00039795D:E10 | CH09LNL   |
| 447    | 376103  | RTA00002676F.g.06.2.P.Seq | F           | M00039295B:D03 | CH09LNL   |
| 448    | 40228   | RTA00002712F.l.18.1.P.Seq | F           | M00027049B:F05 | CH04MAL   |
| 449    | 374606  | RTA00002673F.j.23.1.P.Seq | F           | M00039096A:A05 | CH09LNL   |
| 450    | 378270  | RTA00002680F.h.08.2.P.Seq | F           | M00039801A:H11 | CH09LNL   |
| 451    | 236321  | RTA00002668F.k.14.1.P.Seq | F           | M00033034C:F02 | CH08LNH   |
| 452    | 378676  | RTA00002680F.m.20.2.P.Seq | F           | M00039827B:F07 | CH09LNH   |
| 453    | 373252  | RTA00002670F.k.16.2.P.Seq | F           | M00033451A:H01 | CH09LNL   |
| 454    | 384601  | RTA00002670F.k.06.2.P.Seq | F           | M00033446C:G08 | CH09LNL   |
| 455    | 403772  | RTA00002687.a.03.2.P.Seq  | F           | M00039746C:G08 | CH09LNL   |
| 456    | 379566  | RTA00002683F.k.04.1.P.Seq | F           | M00040081C:E01 | CH09LNL   |
| 457    | 136202  | RTA00002687F.p.05.1.P.Seq | F           | M00040349D:B09 | CH14EDT   |
| 458    | 14317   | RTA00002713F.c.13.1.P.Seq | F           | M00027248A:C02 | CH04MAL   |
| 459    | 374349  | RTA00002672F.j.11.P.Seq   | F           | M00039024B:B10 | CH09LNL   |
| 460    | 403020  | RTA00002687F.a.02.2.P.Seq | F           | M00039746C:A08 | CH14EDT   |
| 461    | 374060  | RTA00002672F.i.07.1.P.Seq | F           | M00039014B:C04 | CH09LNL   |
| 462    | 183399  | RTA00002712F.o.10.1.P.Seq | F           | M00027136C:C09 | CH04MAL   |
| 463    | 373789  | RTA00002671F.c.20.1.P.Seq | F           | M00038259B:G08 | CH09LNL   |
| 464    | 20168   | RTA00002711F.b.22.1.P.Seq | F           | M00022834B:G11 | CH03MAH   |
| 465    | 452641  | RTA00002692F.d.05.2.P.Seq | F           | M00043003C:D08 | CH18CON   |
| 466    | 431370  | RTA00002669F.m.04.2.P.Seq | F           | M00033288B:D12 | CH08LNH   |
| 467    | 153044  | RTA00002713F.j.03.1.P.Seq | F           | M00027476A:C09 | CH04MAL   |
| 468    | 378229  | RTA00002679F.c.16.2.P.Seq | F           | M00039663C:G09 | CH09LNL   |
| 469    | 374328  | RTA00002673F.d.01.1.P.Seq | F           | M00039061B:F08 | CH09LNL   |
| 470    | 39606   | RTA00002713F.i.20.1.P.Seq | F           | M00027468A:C09 | CH04MAL   |
| 471    | 59077   | RTA00002713F.n.01.1.P.Seq | F           | M00027596C:E06 | CH04MAL   |
| 472    | 1935    | RTA00002710F.b.11.1.P.Seq | F           | M00008006B:B03 | CH03MAH   |
| 473    | 379684  | RTA00002681F.c.09.1.P.Seq | F           | M00039851B:G11 | CH09LNL   |
| 474    | 451564  | RTA00002691F.f.12.2.P.Seq | F           | M00043411D:H06 | CH17COHLV |
| 475    | 7571    | RTA00002710F.a.15.1.P.Seq | F           | M00007943D:C09 | CH03MAH   |
| 476    | 129323  | RTA00002713F.k.21.1.P.Seq | F           | M00027525B:D06 | CH04MAL   |
| 477    | 12960   | RTA00002710F.a.23.1.P.Seq | F           | M00007976A:C10 | CH03MAH   |
| 478    | 186730  | RTA00002713F.o.05.1.P.Seq | F           | M00027641C:A03 | CH04MAL   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 479    | 50977   | RTA00002713F.m.24.1.P.Seq | F            | M00027596C:E06 | CH04MAL   |
| 480    | 185884  | RTA00002712F.b.06.1.P.Seq | F            | M00023316C:G08 | CH04MAL   |
| 481    | 19471   | RTA00002708F.g.08.1.P.Seq | F            | M00004197B:H10 | CH01COH   |
| 482    | 45206   | RTA00002710F.c.06.1.P.Seq | F            | M00008063B:A06 | CH03MAH   |
| 483    | 404257  | RTA00002687F.g.06.2.P.Seq | F            | M00040208A:C03 | CH14EDT   |
| 484    | 372997  | RTA00002679F.p.04.1.P.Seq | F            | M00039729A:A10 | CH09LNL   |
| 485    | 43792   | RTA00002713F.k.16.1.P.Seq | F            | M00027520A:C05 | CH04MAL   |
| 486    | 400052  | RTA00002687F.h.13.2.P.Seq | F            | M00040291D:C05 | CH14EDT   |
| 487    | 452194  | RTA00002692F.c.14.2.P.Seq | F            | M00042988A:F06 | CH18CON   |
| 488    | 24034   | RTA00002710F.b.06.1.P.Seq | F            | M00007992C:F06 | CH03MAH   |
| 489    | 447544  | RTA00002689F.e.18.1.P.Seq | F            | M00042905D:D02 | CH15CON   |
| 490    | 401872  | RTA00002686F.c.23.1.P.Seq | F            | M00040141D:F05 | CH13EDT   |
| 491    | 376553  | RTA00002674F.g.19.2.P.Seq | F            | M00039139A:C09 | CH09LNL   |
| 492    | 455051  | RTA00002694F.a.07.1.P.Seq | F            | M00042595A:A11 | CH20COHLV |
| 493    | 16760   | RTA00002708F.j.03.1.P.Seq | F            | M00004393B:E07 | CH01COH   |
| 494    | 374174  | RTA00002672F.i.12.2.P.Seq | F            | M00039015A:D07 | CH09LNL   |
| 495    | 374283  | RTA00002672F.k.21.2.P.Seq | F            | M00039030B:E02 | CH09LNL   |
| 496    | 375772  | RTA00002681F.o.24.1.P.Seq | F            | M00039909C:G05 | CH09LNL   |
| 497    | 376417  | RTA00002678F.i.03.2.P.Seq | F            | M00039477D:A10 | CH09LNL   |
| 498    | 428971  | RTA00002666F.o.02.1.P.Seq | F            | M00032678C:D06 | CH08LNH   |
| 499    | 394098  | RTA00002681F.j.15.1.P.Seq | F            | M00039887C:E07 | CH09LNL   |
| 500    | 379761  | RTA00002670F.n.03.1.P.Seq | F            | M00033561C:A02 | CH09LNL   |
| 501    | 374266  | RTA00002674F.i.08.1.P.Seq | F            | M00039144C:E06 | CH09LNL   |
| 502    | 372946  | RTA00002670F.l.07.1.P.Seq | F            | M00033457D:A05 | CH09LNL   |
| 503    | 228909  | RTA00002664F.e.08.1.P.Seq | F            | M00027085C:E11 | CH04MAL   |
| 504    | 427524  | RTA00002665F.e.05.1.P.Seq | F            | M00028354D:A03 | CH08LNH   |
| 505    | 380413  | RTA00002680F.k.19.2.P.Seq | F            | M00039816C:D05 | CH09LNL   |
| 506    | 373866  | RTA00002671F.c.24.2.P.Seq | F            | M00038259C:H09 | CH09LNL   |
| 507    | 427202  | RTA00002665F.g.15.1.P.Seq | F            | M00028617C:A12 | CH08LNH   |
| 508    | 373000  | RTA00002670F.j.13.1.P.Seq | F            | M00033437C:C03 | CH09LNL   |
| 509    | 378838  | RTA00002678F.p.11.2.P.Seq | F            | M00039637C:A10 | CH09LNL   |
| 510    | 24945   | RTA00002710F.p.05.1.P.Seq | F            | M00022739A:B03 | CH03MAH   |
| 511    | 20277   | RTA00002710F.e.17.1.P.Seq | F            | M00021972D:C11 | CH03MAH   |
| 512    | 20820   | RTA00002710F.e.02.1.P.Seq | F            | M00021919C:A10 | CH03MAH   |
| 513    | 376791  | RTA00002674F.l.17.2.P.Seq | F            | M00039166B:G06 | CH09LNL   |
| 514    | 9809    | RTA00002710F.g.12.1.P.Seq | F            | M00022178B:D06 | CH03MAH   |
| 515    | 429562  | RTA00002667F.m.03.1.P.Seq | F            | M00032853D:G12 | CH08LNH   |
| 516    | 12920   | RTA00002710F.e.15.1.P.Seq | F            | M00021964C:E10 | CH03MAH   |
| 517    | 377565  | RTA00002684F.h.19.1.P.Seq | F            | M00040309A:E11 | CH09LNL   |
| 518    | 429356  | RTA00002668F.d.23.1.P.Seq | F            | M00032933A:C10 | CH08LNH   |
| 519    | 427634  | RTA00002665F.f.09.1.P.Seq | F            | M00028369D:E08 | CH08LNH   |
| 520    | 427713  | RTA00002665F.e.23.1.P.Seq | F            | M00028364C:G08 | CH08LNH   |
| 521    | 373607  | RTA00002674F.d.15.2.P.Seq | F            | M00039127D:E10 | CH09LNL   |
| 522    | 378781  | RTA00002674F.o.14.1.P.Seq | F            | M00039196B:H06 | CH09LNL   |
| 523    | 429361  | RTA00002666F.d.11.1.P.Seq | F            | M00032550D:C02 | CH08LNH   |
| 524    | 126754  | RTA00002663F.a.16.1.P.Seq | F            | M00008045A:H02 | CH03MAH   |
| 525    | 428047  | RTA00002665F.k.10.1.P.Seq | F            | M00031417C:G09 | CH08LNH   |
| 526    | 18863   | RTA00002709F.d.15.1.P.Seq | F            | M00005625A:C02 | CH02COH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY  |
|--------|---------|---------------------------|-------------|----------------|----------|
| 527    | 379761  | RTA00002670F.n.03.2.P.Seq | F           | M00033561C:A02 | CH09LNL  |
| 528    | 46407   | RTA00002665F.c.10.1.P.Seq | F           | M00028196D:A03 | CH08LNLH |
| 529    | 21365   | RTA00002709F.k.06.1.P.Seq | F           | M00007012D:H08 | CH02COH  |
| 530    | 427466  | RTA00002665F.b.11.1.P.Seq | F           | M00028184D:G10 | CH08LNLH |
| 531    | 400265  | RTA00002685F.c.03.2.P.Seq | F           | M00039374B:B07 | CH12EDT  |
| 532    | 380056  | RTA00002680F.a.16.2.P.Seq | F           | M00039773D:F11 | CH09LNL  |
| 533    | 375324  | RTA00002678F.l.12.2.P.Seq | F           | M00039612B:B10 | CH09LNL  |
| 534    | 25165   | RTA00002710F.k.17.1.P.Seq | F           | M00022496B:E12 | CH03MAH  |
| 535    | 401296  | RTA00002685F.h.23.2.P.Seq | F           | M00039529C:D07 | CH12EDT  |
| 536    | 394098  | RTA00002681F.j.15.2.P.Seq | F           | M00039887C:E07 | CH09LNL  |
| 537    | 17430   | RTA00002710F.i.11.1.P.Seq | F           | M00022365D:A03 | CH03MAH  |
| 538    | 373820  | RTA00002674F.d.06.1.P.Seq | F           | M00039127A:G11 | CH09LNL  |
| 539    | 378548  | RTA00002672F.g.14.2.P.Seq | F           | M00039004B:C11 | CH09LNL  |
| 540    | 222679  | RTA00002664F.f.18.2.P.Seq | F           | M00027228D:A01 | CH04MAL  |
| 541    | 376874  | RTA00002670F.e.23.2.P.Seq | F           | M00033375A:G04 | CH09LNL  |
| 542    | 21329   | RTA00002709F.b.08.1.P.Seq | F           | M00005379A:E04 | CH02COH  |
| 543    | 119905  | RTA00002710F.p.13.1.P.Seq | F           | M00022785C:G06 | CH03MAH  |
| 544    | 377028  | RTA00002678F.n.21.2.P.Seq | F           | M00039631A:C10 | CH09LNL  |
| 545    | 373351  | RTA00002671F.l.18.3.P.Seq | F           | M00038327D:A05 | CH09LNL  |
| 546    | 376082  | RTA00002674F.m.17.1.P.Seq | F           | M00039171B:D11 | CH09LNL  |
| 547    | 376987  | RTA00002678F.g.21.2.P.Seq | F           | M00039472C:B08 | CH09LNL  |
| 548    | 61921   | RTA00002661F.g.08.1.P.Seq | F           | M00003995B:E03 | CH01COH  |
| 549    | 373486  | RTA00002672F.b.03.2.P.Seq | F           | M00038635B:C08 | CH09LNL  |
| 550    | 380355  | RTA00002670F.o.06.2.P.Seq | F           | M00033570C:C10 | CH09LNL  |
| 551    | 430295  | RTA00002667F.h.14.1.P.Seq | F           | M00032808B:G10 | CH08LNLH |
| 552    | 379221  | RTA00002682F.n.01.1.P.Seq | F           | M00040017D:G03 | CH09LNL  |
| 553    | 373532  | RTA00002672F.d.10.2.P.Seq | F           | M00038991A:D01 | CH09LNL  |
| 554    | 375633  | RTA00002677F.m.05.2.P.Seq | F           | M00039417B:F01 | CH09LNL  |
| 555    | 378356  | RTA00002681F.f.07.1.P.Seq | F           | M00039866B:A08 | CH09LNL  |
| 556    | 376196  | RTA00002674F.m.12.1.P.Seq | F           | M00039170C:F05 | CH09LNL  |
| 557    | 375115  | RTA00002673F.d.24.2.P.Seq | F           | M00039066D:G08 | CH09LNL  |
| 558    | 375115  | RTA00002673F.e.01.2.P.Seq | F           | M00039066D:G08 | CH09LNL  |
| 559    | 378600  | RTA00002679F.i.03.1.P.Seq | F           | M00039686C:E06 | CH09LNL  |
| 560    | 375351  | RTA00002680F.e.15.1.P.Seq | F           | M00039792A:B04 | CH09LNL  |
| 561    | 25237   | RTA00002710F.n.23.1.P.Seq | F           | M00022671B:A08 | CH03MAH  |
| 562    | 193503  | RTA00002663F.n.15.1.P.Seq | F           | M00023039D:B05 | CH03MAH  |
| 563    | 428268  | RTA00002667F.b.01.1.P.Seq | F           | M00032724A:C05 | CH08LNLH |
| 564    | 379440  | RTA00002683F.j.21.2.P.Seq | F           | M00040080C:C06 | CH09LNL  |
| 565    | 374502  | RTA00002673F.i.08.2.P.Seq | F           | M00039080C:H06 | CH09LNL  |
| 566    | 240615  | RTA00002672F.e.19.2.P.Seq | F           | M00038995D:E05 | CH09LNL  |
| 567    | 379207  | RTA00002670F.b.07.2.P.Seq | F           | M00033306D:G08 | CH09LNL  |
| 568    | 427893  | RTA00002665F.k.19.1.P.Seq | F           | M00031419D:C04 | CH08LNLH |
| 569    | 377530  | RTA00002684F.g.19.2.P.Seq | F           | M00040305A:D11 | CH09LNL  |
| 570    | 429707  | RTA00002668F.c.11.1.P.Seq | F           | M00032918C:B10 | CH08LNLH |
| 571    | 427610  | RTA00002665F.i.04.1.P.Seq | F           | M00028770A:D04 | CH08LNLH |
| 572    | 100699  | RTA00002662F.b.22.2.P.Seq | F           | M00006680B:D02 | CH02COH  |
| 573    | 378974  | RTA00002682F.m.21.1.P.Seq | F           | M00040017A:C06 | CH09LNL  |
| 574    | 373607  | RTA00002674F.d.15.1.P.Seq | F           | M00039127D:E10 | CH09LNL  |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|--------------|----------------|---------|
| 575    | 262951  | RTA00002665F.d.04.3.P.Seq | F            | M00028215D:F03 | CH08LNH |
| 576    | 30748   | RTA00002713F.e.11.1.P.Seq | F            | M00027301B:B08 | CH04MAL |
| 577    | 16116   | RTA00002714F.c.11.1.P.Seq | F            | M00027837C:D09 | CH04MAL |
| 578    | 379211  | RTA00002682F.p.20.1.P.Seq | F            | M00040029A:G04 | CH09LNL |
| 579    | 430689  | RTA00002669F.i.24.1.P.Seq | F            | M00033243B:A05 | CH08LNH |
| 580    | 374122  | RTA00002673F.l.22.2.P.Seq | F            | M00039104D:C09 | CH09LNL |
| 581    | 376521  | RTA00002677F.h.06.2.P.Seq | F            | M00039398A:B10 | CH09LNL |
| 582    | 372834  | RTA00002670F.b.12.2.P.Seq | F            | M00033308B:G05 | CH09LNL |
| 583    | 379014  | RTA00002682F.o.02.1.P.Seq | F            | M00040022C:D06 | CH09LNL |
| 584    | 376344  | RTA00002677F.b.18.2.P.Seq | F            | M00039340B:G08 | CH09LNL |
| 585    | 376485  | RTA00002676F.f.01.2.P.Seq | F            | M00039288C:B11 | CH09LNL |
| 586    | 21661   | RTA00002709F.e.18.1.P.Seq | F            | M00005820C:E04 | CH02COH |
| 587    | 376539  | RTA00002675F.b.18.1.P.Seq | F            | M00039211A:C12 | CH09LNL |
| 588    | 431645  | RTA00002669F.h.15.3.P.Seq | F            | M00033223B:H07 | CH08LNH |
| 589    | 163293  | RTA00002714F.c.20.1.P.Seq | F            | M00028120D:F12 | CH04MAL |
| 590    | 178614  | RTA00002713F.c.20.1.P.Seq | F            | M00027263A:F10 | CH04MAL |
| 591    | 373274  | RTA00002670F.i.22.2.P.Seq | F            | M00033432B:H10 | CH09LNL |
| 592    | 379820  | RTA00002679F.f.15.1.P.Seq | F            | M00039677A:B08 | CH09LNL |
| 593    | 160536  | RTA00002663F.f.10.1.P.Seq | F            | M00022233C:A12 | CH03MAH |
| 594    | 373313  | RTA00002671F.m.02.1.P.Seq | F            | M00038328D:A03 | CH09LNL |
| 595    | 26429   | RTA00002712F.k.23.1.P.Seq | F            | M00027022D:G11 | CH04MAL |
| 596    | 17983   | RTA00002711F.f.10.1.P.Seq | F            | M00022979A:D05 | CH03MAH |
| 597    | 375388  | RTA00002681F.j.22.2.P.Seq | F            | M00039888B:D03 | CH09LNL |
| 598    | 63005   | RTA00002712F.m.21.1.P.Seq | F            | M00027094A:B03 | CH04MAL |
| 599    | 23030   | RTA00002709F.b.10.2.P.Seq | F            | M00005384A:C11 | CH02COH |
| 600    | 372946  | RTA00002670F.l.07.2.P.Seq | F            | M00033457D:A05 | CH09LNL |
| 601    | 375351  | RTA00002680F.e.15.2.P.Seq | F            | M00039792A:B04 | CH09LNL |
| 602    | 374502  | RTA00002673F.i.08.1.P.Seq | F            | M00039080C:H06 | CH09LNL |
| 603    | 376911  | RTA00002682F.e.09.1.P.Seq | F            | M00039938C:A08 | CH09LNL |
| 604    | 376024  | RTA00002675F.n.15.1.P.Seq | F            | M00039257D:C03 | CH09LNL |
| 605    | 377194  | RTA00002679F.h.20.1.P.Seq | F            | M00039685A:A08 | CH09LNL |
| 606    | 379643  | RTA00002682F.g.08.1.P.Seq | F            | M00039978A:G03 | CH09LNL |
| 607    | 379610  | RTA00002680F.k.11.1.P.Seq | F            | M00039815C:F09 | CH09LNL |
| 608    | 25613   | RTA00002711F.g.06.1.P.Seq | F            | M00023024D:F12 | CH03MAH |
| 609    | 207466  | RTA00002664F.j.08.1.P.Seq | F            | M00027733A:A02 | CH04MAL |
| 610    | 400052  | RTA00002687F.h.13.1.P.Seq | F            | M00040291D:C05 | CH14EDT |
| 611    | 21290   | RTA00002712F.g.01.1.P.Seq | F            | M00026859D:D01 | CH04MAL |
| 612    | 375975  | RTA00002675F.n.18.1.P.Seq | F            | M00039258D:B08 | CH09LNL |
| 613    | 46804   | RTA00002712F.n.19.1.P.Seq | F            | M00027121D:C05 | CH04MAL |
| 614    | 69863   | RTA00002712F.i.18.1.P.Seq | F            | M00026935C:B04 | CH04MAL |
| 615    | 375285  | RTA00002676F.g.18.2.P.Seq | F            | M00039298B:B06 | CH09LNL |
| 616    | 373000  | RTA00002670F.j.13.2.P.Seq | F            | M00033437C:C03 | CH09LNL |
| 617    | 378679  | RTA00002681F.f.16.1.P.Seq | F            | M00039869B:F06 | CH09LNL |
| 618    | 45407   | RTA00002712F.k.11.1.P.Seq | F            | M00027016A:B06 | CH04MAL |
| 619    | 16838   | RTA00002712F.e.23.1.P.Seq | F            | M00026803A:F08 | CH04MAL |
| 620    | 186425  | RTA00002713F.c.04.1.P.Seq | F            | M00027236A:E04 | CH04MAL |
| 621    | 376485  | RTA00002676F.e.24.2.P.Seq | F            | M00039288C:B11 | CH09LNL |
| 622    | 41108   | RTA00002712F.n.12.1.P.Seq | F            | M00027108C:B03 | CH04MAL |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 623    | 430876  | RTA00002669F.c.02.1.P.Seq | F           | M00033186C:D11 | CH08LNH   |
| 624    | 185716  | RTA00002713F.l.07.1.P.Seq | F           | M00027537C:B01 | CH04MAL   |
| 625    | 85338   | RTA00002712F.b.18.1.P.Seq | F           | M00023333D:C12 | CH04MAL   |
| 626    | 185597  | RTA00002713F.m.23.1.P.Seq | F           | M00027596A:A10 | CH04MAL   |
| 627    | 139348  | RTA00002713F.k.23.1.P.Seq | F           | M00027526D:F03 | CH04MAL   |
| 628    | 454665  | RTA00002693F.d.15.2.P.Seq | F           | M00043164C:E12 | CH19COP   |
| 629    | 186387  | RTA00002713F.l.01.1.P.Seq | F           | M00027528C:B10 | CH04MAL   |
| 630    | 186387  | RTA00002713F.k.24.1.P.Seq | F           | M00027528C:B10 | CH04MAL   |
| 631    | 21093   | RTA00002708F.h.20.1.P.Seq | F           | M00004308C:C06 | CH01COH   |
| 632    | 20827   | RTA00002710F.c.23.1.P.Seq | F           | M00021671D:F12 | CH03MAH   |
| 633    | 21290   | RTA00002712F.f.24.1.P.Seq | F           | M00026859D:D01 | CH04MAL   |
| 634    | 17646   | RTA00002710F.d.22.1.P.Seq | F           | M00021908D:G12 | CH03MAH   |
| 635    | 402817  | RTA00002686F.a.10.1.P.Seq | F           | M00039736D:G08 | CH13EDT   |
| 636    | 42854   | RTA00002713F.n.09.1.P.Seq | F           | M00027615A:F10 | CH04MAL   |
| 637    | 430876  | RTA00002669F.c.02.3.P.Seq | F           | M00033186C:D11 | CH08LNH   |
| 638    | 378641  | RTA00002679F.a.21.2.P.Seq | F           | M00039655C:E08 | CH09LNL   |
| 639    | 375848  | RTA00002674F.m.03.2.P.Seq | F           | M00039168C:A04 | CH09LNL   |
| 640    | 36165   | RTA00002708F.i.06.1.P.Seq | F           | M00004340C:C07 | CH01COH   |
| 641    | 456506  | RTA00002694F.d.05.1.P.Seq | F           | M00043492A:E01 | CH20COHLV |
| 642    | 374450  | RTA00002672F.i.05.2.P.Seq | F           | M00039014A:H10 | CH09LNL   |
| 643    | 378949  | RTA00002683F.o.21.2.P.Seq | F           | M00040100D:B06 | CH09LNL   |
| 644    | 373313  | RTA00002671F.m.02.2.P.Seq | F           | M00038328D:A03 | CH09LNL   |
| 645    | 377861  | RTA00002681F.m.20.1.P.Seq | F           | M00039898A:A08 | CH09LNL   |
| 646    | 431196  | RTA00002669F.f.07.2.P.Seq | F           | M00033204B:A07 | CH08LNH   |
| 647    | 372795  | RTA00002683F.a.06.1.P.Seq | F           | M00040032A:B03 | CH09LNL   |
| 648    | 42340   | RTA00002661F.b.03.1.P.Seq | F           | M00001439C:H06 | CH01COH   |
| 649    | 374410  | RTA00002674F.k.11.1.P.Seq | F           | M00039158B:G12 | CH09LNL   |
| 650    | 374623  | RTA00002674F.a.01.2.P.Seq | F           | M00039118D:A06 | CH09LNL   |
| 651    | 431612  | RTA00002669F.e.23.2.P.Seq | F           | M00033202D:G06 | CH08LNH   |
| 652    | 240615  | RTA00002672F.e.19.1.P.Seq | F           | M00038995D:E05 | CH09LNL   |
| 653    | 428508  | RTA00002666F.d.01.1.P.Seq | F           | M00032545B:H09 | CH08LNH   |
| 654    | 235780  | RTA00002666F.d.03.1.P.Seq | F           | M00032545D:G05 | CH08LNH   |
| 655    | 17890   | RTA00002710F.e.11.1.P.Seq | F           | M00021955A:H02 | CH03MAH   |
| 656    | 20100   | RTA00002710F.g.11.1.P.Seq | F           | M00022175D:D12 | CH03MAH   |
| 657    | 4458    | RTA00002710F.g.18.1.P.Seq | F           | M00022184C:C11 | CH03MAH   |
| 658    | 378347  | RTA00002681F.h.07.2.P.Seq | F           | M00039875D:A10 | CH09LNL   |
| 659    | 373477  | RTA00002672F.b.23.1.P.Seq | F           | M00038639B:C03 | CH09LNL   |
| 660    | 15596   | RTA00002710F.g.02.1.P.Seq | F           | M00022135C:B05 | CH03MAH   |
| 661    | 21028   | RTA00002709F.l.09.1.P.Seq | F           | M00007108B:A02 | CH02COH   |
| 662    | 374063  | RTA00002672F.h.15.2.P.Seq | F           | M00039011D:C10 | CH09LNL   |
| 663    | 380686  | RTA00002684F.a.03.2.P.Seq | F           | M00040107B:H07 | CH09LNL   |
| 664    | 402950  | RTA00002686F.g.11.1.P.Seq | F           | M00040181B:H09 | CH13EDT   |
| 665    | 428064  | RTA00002665F.l.04.1.P.Seq | F           | M00031485D:G02 | CH08LNH   |
| 666    | 23310   | RTA00002708F.e.10.1.P.Seq | F           | M00004046C:A08 | CH01COH   |
| 667    | 376233  | RTA00002677F.b.15.2.P.Seq | F           | M00039339C:F03 | CH09LNL   |
| 668    | 375848  | RTA00002674F.m.03.1.P.Seq | F           | M00039168C:A04 | CH09LNL   |
| 669    | 242251  | RTA00002665F.i.08.1.P.Seq | F           | M00028772C:B09 | CH08LNH   |
| 670    | 374064  | RTA00002672F.f.15.2.P.Seq | F           | M00038999D:C11 | CH09LNL   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|-------------|-----------------|---------|
| 671    | 146260  | RTA00002663F.d.17.1.P.Seq | F           | M00022099B:D06  | CH03MAH |
| 672    | 375575  | RTA00002677F.e.22.2.P.Seq | F           | M00039385B:E09  | CH09LNL |
| 673    | 355518  | RTA00002665F.c.15.3.P.Seq | F           | M00028201B:H12  | CH08LNL |
| 674    | 184223  | RTA00002662F.b.08.2.P.Seq | F           | M00005539D:G01  | CH02COH |
| 675    | 213306  | RTA00002664F.e.07.2.P.Seq | F           | M00027078A:B02  | CH04MAL |
| 676    | 429566  | RTA00002668F.b.04.1.P.Seq | F           | M00032907A:G04  | CH08LNL |
| 677    | 378656  | RTA00002682F.c.09.1.P.Seq | F           | M00039927A:F04  | CH09LNL |
| 678    | 427760  | RTA00002668F.e.23.1.P.Seq | F           | M00032940A:C02  | CH08LNL |
| 679    | 372795  | RTA00002683F.a.06.2.P.Seq | F           | M00040032A:B03  | CH09LNL |
| 680    | 429340  | RTA00002666F.f.12.1.P.Seq | F           | M00032577A:C04  | CH08LNL |
| 681    | 429822  | RTA00002668F.e.17.1.P.Seq | F           | M00032939B:E07  | CH08LNL |
| 682    | 375224  | RTA00002680F.d.22.2.P.Seq | F           | M00039788B:A06  | CH09LNL |
| 683    | 378347  | RTA00002681F.h.07.1.P.Seq | F           | M00039875D:A10  | CH09LNL |
| 684    | 380109  | RTA00002682F.i.17.1.P.Seq | F           | M00039987C:G089 | CH09LNL |
| 685    | 379001  | RTA00002683F.o.02.1.P.Seq | F           | M00040097A:C12  | CH09LNL |
| 686    | 375348  | RTA00002676F.i.12.3.P.Seq | F           | M00039304D:B09  | CH09LNL |
| 687    | 377889  | RTA00002672F.c.08.2.P.Seq | F           | M00038661A:A07  | CH09LNL |
| 688    | 429883  | RTA00002667F.g.05.1.P.Seq | F           | M00032793A:F06  | CH08LNL |
| 689    | 377067  | RTA00002682F.l.24.1.P.Seq | F           | M00040014B:D01  | CH09LNL |
| 690    | 378001  | RTA00002681F.m.22.2.P.Seq | F           | M00039898D:C06  | CH09LNL |
| 691    | 45298   | RTA00002710F.j.21.1.P.Seq | F           | M00022433A:E02  | CH03MAH |
| 692    | 375431  | RTA00002680F.f.03.1.P.Seq | F           | M00039793D:C05  | CH09LNL |
| 693    | 377861  | RTA00002681F.m.20.2.P.Seq | F           | M00039898A:A08  | CH09LNL |
| 694    | 428610  | RTA00002667F.e.09.1.P.Seq | F           | M00032766C:A04  | CH08LNL |
| 695    | 20765   | RTA00002710F.i.10.1.P.Seq | F           | M00022363C:G12  | CH03MAH |
| 696    | 27601   | RTA00002713F.e.23.1.P.Seq | F           | M00027314C:D09  | CH04MAL |
| 697    | 430540  | RTA00002668F.o.20.2.P.Seq | F           | M00033140D:F06  | CH08LNL |
| 698    | 381024  | RTA00002670F.h.23.2.P.Seq | F           | M00033424B:A04  | CH09LNL |
| 699    | 16454   | RTA00002709F.f.07.1.P.Seq | F           | M00006599D:B02  | CH02COH |
| 700    | 372898  | RTA00002670F.i.03.2.P.Seq | F           | M00033424D:H12  | CH09LNL |
| 701    | 373681  | RTA00002671F.d.20.1.P.Seq | F           | M00038272D:F11  | CH09LNL |
| 702    | 82260   | RTA00002684F.h.06.2.P.Seq | F           | M00040307B:F01  | CH09LNL |
| 703    | 377343  | RTA00002684F.g.04.2.P.Seq | F           | M00040302C:A04  | CH09LNL |
| 704    | 374747  | RTA00002676F.e.07.2.P.Seq | F           | M00039286A:C06  | CH09LNL |
| 705    | 185848  | RTA00002712F.m.11.1.P.Seq | F           | M00027080A:B01  | CH04MAL |
| 706    | 374311  | RTA00002676F.e.18.2.P.Seq | F           | M00039287C:A06  | CH09LNL |
| 707    | 278923  | RTA00002667F.b.10.1.P.Seq | F           | M00032726C:C01  | CH08LNL |
| 708    | 378667  | RTA00002681F.b.11.2.P.Seq | F           | M00039847A:F06  | CH09LNL |
| 709    | 380454  | RTA00002673F.j.16.1.P.Seq | F           | M00039084D:D07  | CH09LNL |
| 710    | 381576  | RTA00002670F.i.04.2.P.Seq | F           | M00033425A:C10  | CH09LNL |
| 711    | 375067  | RTA00002675F.o.03.1.P.Seq | F           | M00039260C:G03  | CH09LNL |
| 712    | 89706   | RTA00002714F.a.11.1.P.Seq | F           | M00027741B:F09  | CH04MAL |
| 713    | 10583   | RTA00002711F.h.11.1.P.Seq | F           | M00023100A:E12  | CH03MAH |
| 714    | 379982  | RTA00002682F.i.16.1.P.Seq | F           | M00039987C:E12  | CH09LNL |
| 715    | 378532  | RTA00002680F.n.04.3.P.Seq | F           | M00039828B:C05  | CH09LNL |
| 716    | 379776  | RTA00002680F.a.22.2.P.Seq | F           | M00039774C:A03  | CH09LNL |
| 717    | 374136  | RTA00002673F.f.16.1.P.Seq | F           | M00039072C:C03  | CH09LNL |
| 718    | 98471   | RTA00002663F.j.21.1.P.Seq | F           | M00022670D:H11  | CH03MAH |



| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 719    | 125365  | RTA00002668F.j.07.1.P.Seq | F            | M00033019B:E10 | CH08LNH   |
| 720    | 375431  | RTA00002680F.f.03.2.P.Seq | F            | M00039793D:C05 | CH09LNL   |
| 721    | 62826   | RTA00002661F.g.20.1.P.Seq | F            | M00004105D:D05 | CH01COH   |
| 722    | 379972  | RTA00002679F.e.10.1.P.Seq | F            | M00039672D:D10 | CH09LNL   |
| 723    | 377554  | RTA00002679F.f.10.1.P.Seq | F            | M00039675D:B03 | CH09LNL   |
| 724    | 230479  | RTA00002664F.c.16.2.P.Seq | F            | M00026915B:C06 | CH04MAL   |
| 725    | 98872   | RTA00002663F.j.19.1.P.Seq | F            | M00022668B:B12 | CH03MAH   |
| 726    | 42635   | RTA00002679F.h.18.1.P.Seq | F            | M00039684D:B08 | CH09LNL   |
| 727    | 379044  | RTA00002679F.a.10.2.P.Seq | F            | M00039652B:D05 | CH09LNL   |
| 728    | 96093   | RTA00002663F.j.07.1.P.Seq | F            | M00022640C:C12 | CH03MAH   |
| 729    | 403642  | RTA00002687F.d.01.2.P.Seq | F            | M00039945C:F09 | CH14EDT   |
| 730    | 400921  | RTA00002685F.b.18.2.P.Seq | F            | M00039371B:H06 | CH12EDT   |
| 731    | 93587   | RTA00002663F.k.10.1.P.Seq | F            | M00022731A:D02 | CH03MAH   |
| 732    | 79951   | RTA00002713F.c.18.1.P.Seq | F            | M00027258A:A07 | CH04MAL   |
| 733    | 176509  | RTA00002686F.b.09.1.P.Seq | F            | M00039756B:H06 | CH13EDT   |
| 734    | 451753  | RTA00002694F.e.06.1.P.Seq | F            | M00043634A:C10 | CH20COHLV |
| 735    | 186266  | RTA00002713F.c.16.1.P.Seq | F            | M00027256B:H09 | CH04MAL   |
| 736    | 235052  | RTA00002692F.a.15.2.P.Seq | F            | M00042626B:D08 | CH18CON   |
| 737    | 377233  | RTA00002682F.e.23.1.P.Seq | F            | M00039940D:G08 | CH09LNL   |
| 738    | 378532  | RTA00002680F.n.04.2.P.Seq | F            | M00039828B:C05 | CH09LNL   |
| 739    | 177932  | RTA00002713F.b.22.1.P.Seq | F            | M00027233B:C01 | CH04MAL   |
| 740    | 9332    | RTA00002712F.p.18.1.P.Seq | F            | M00027179D:E06 | CH04MAL   |
| 741    | 240318  | RTA00002687F.d.04.2.P.Seq | F            | M00039947A:D06 | CH14EDT   |
| 742    | 404260  | RTA00002687F.c.11.2.P.Seq | F            | M00039942D:C01 | CH14EDT   |
| 743    | 93767   | RTA00002712F.g.09.1.P.Seq | F            | M00026868C:E11 | CH04MAL   |
| 744    | 185642  | RTA00002712F.f.20.1.P.Seq | F            | M00026856D:F02 | CH04MAL   |
| 745    | 447544  | RTA00002689F.e.18.3.P.Seq | F            | M00042905D:D02 | CH15CON   |
| 746    | 403274  | RTA00002687F.b.10.2.P.Seq | F            | M00039766A:G07 | CH14EDT   |
| 747    | 404257  | RTA00002687F.g.06.1.P.Seq | F            | M00040208A:C03 | CH14EDT   |
| 748    | 403868  | RTA00002687F.k.05.2.P.Seq | F            | M00040318C:H11 | CH14EDT   |
| 749    | 450074  | RTA00002691F.e.12.2.P.Seq | F            | M00043392D:C11 | CH17COHLV |
| 750    | 404520  | RTA00002687F.f.05.2.P.Seq | F            | M00040202A:F05 | CH14EDT   |
| 751    | 451789  | RTA00002692F.b.04.2.P.Seq | F            | M00042956C:B06 | CH18CON   |
| 752    | 455178  | RTA00002694F.b.19.1.P.Seq | F            | M00043447A:C07 | CH20COHLV |
| 753    | 455136  | RTA00002694F.a.08.1.P.Seq | F            | M00042595A:B01 | CH20COHLV |
| 754    | 379001  | RTA00002683F.o.02.2.P.Seq | F            | M00040097A:C12 | CH09LNL   |
| 755    | 374763  | RTA00002673F.p.21.1.P.Seq | F            | M00039118B:C05 | CH09LNL   |
| 756    | 402508  | RTA00002686F.o.15.1.P.Seq | F            | M00040281D:B01 | CH13EDT   |
| 757    | 431370  | RTA00002669F.m.04.3.P.Seq | F            | M00033288B:D12 | CH08LNH   |
| 758    | 380500  | RTA00002670F.p.19.1.P.Seq | F            | M00033583B:E06 | CH09LNL   |
| 759    | 376743  | RTA00002678F.e.22.2.P.Seq | F            | M00039461A:F04 | CH09LNL   |
| 760    | 191690  | RTA00002673F.m.19.1.P.Seq | F            | M00039107C:E04 | CH09LNL   |
| 761    | 374264  | RTA00002671F.p.21.2.P.Seq | F            | M00038620B:E09 | CH09LNL   |
| 762    | 373020  | RTA00002671F.b.20.2.P.Seq | F            | M00033595A:C11 | CH09LNL   |
| 763    | 375231  | RTA00002671F.m.20.2.P.Seq | F            | M00038387B:A07 | CH09LNL   |
| 764    | 16180   | RTA00002709F.j.17.1.P.Seq | F            | M00006977D:A03 | CH02COH   |
| 765    | 379403  | RTA00002683F.c.17.2.P.Seq | F            | M00040041C:C09 | CH09LNL   |
| 766    | 375382  | RTA00002677F.d.24.2.P.Seq | F            | M00039381D:C02 | CH09LNL   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|-------------|----------------|---------|
| 767    | 379653  | RTA00002683F.c.03.2.P.Seq | F           | M00040038D:G04 | CH09LNL |
| 768    | 377858  | RTA00002681F.e.14.2.P.Seq | F           | M00039864A:A07 | CH09LNL |
| 769    | 430861  | RTA00002668F.h.18.1.P.Seq | F           | M00032995C:C05 | CH08LNH |
| 770    | 376128  | RTA00002677F.a.11.2.P.Seq | F           | M00039334B:E03 | CH09LNL |
| 771    | 375009  | RTA00002676F.n.20.2.P.Seq | F           | M00039322A:F04 | CH09LNL |
| 772    | 429816  | RTA00002667F.n.22.1.P.Seq | F           | M00032871D:E11 | CH08LNH |
| 773    | 375657  | RTA00002681F.h.13.2.P.Seq | F           | M00039877C:C03 | CH09LNL |
| 774    | 427889  | RTA00002666F.b.14.1.P.Seq | F           | M00032530D:C02 | CH08LNH |
| 775    | 376761  | RTA00002677F.g.03.2.P.Seq | F           | M00039391D:F08 | CH09LNL |
| 776    | 44025   | RTA00002684F.b.24.2.P.Seq | F           | M00040115B:A04 | CH09LNL |
| 777    | 44025   | RTA00002684F.c.01.2.P.Seq | F           | M00040115B:A04 | CH09LNL |
| 778    | 392524  | RTA00002681F.p.04.2.P.Seq | F           | M00039909D:C02 | CH09LNL |
| 779    | 427252  | RTA00002665F.b.13.1.P.Seq | F           | M00028185B:A06 | CH08LNH |
| 780    | 374927  | RTA00002673F.e.12.1.P.Seq | F           | M00039068C:E06 | CH09LNL |
| 781    | 378226  | RTA00002680F.g.09.1.P.Seq | F           | M00039797C:G05 | CH09LNL |
| 782    | 217964  | RTA00002664F.g.08.2.P.Seq | F           | M00027299B:B12 | CH04MAL |
| 783    | 376368  | RTA00002677F.b.14.2.P.Seq | F           | M00039339A:H07 | CH09LNL |
| 784    | 377719  | RTA00002677F.j.11.2.P.Seq | F           | M00039407B:G02 | CH09LNL |
| 785    | 378081  | RTA00002677F.e.16.2.P.Seq | F           | M00039384C:E02 | CH09LNL |
| 786    | 89267   | RTA00002662F.b.01.2.P.Seq | F           | M00005445D:B01 | CH02COH |
| 787    | 374927  | RTA00002673F.e.12.2.P.Seq | F           | M00039068C:E06 | CH09LNL |
| 788    | 279054  | RTA00002667F.b.23.1.P.Seq | F           | M00032731B:C10 | CH08LNH |
| 789    | 377283  | RTA00002682F.m.19.1.P.Seq | F           | M00040016C:H12 | CH09LNL |
| 790    | 45318   | RTA00002710F.l.05.1.P.Seq | F           | M00022533A:A08 | CH03MAH |
| 791    | 188292  | RTA00002664F.e.23.2.P.Seq | F           | M00027162B:F05 | CH04MAL |
| 792    | 378872  | RTA00002683F.c.20.2.P.Seq | F           | M00040042B:A10 | CH09LNL |
| 793    | 427252  | RTA00002665F.b.13.3.P.Seq | F           | M00028185B:A06 | CH08LNH |
| 794    | 380618  | RTA00002673F.j.12.2.P.Seq | F           | M00039084C:G07 | CH09LNL |
| 795    | 35646   | RTA00002667F.g.16.1.P.Seq | F           | M00032797B:G02 | CH08LNH |
| 796    | 46407   | RTA00002665F.c.10.3.P.Seq | F           | M00028196D:A03 | CH08LNH |
| 797    | 373720  | RTA00002674F.c.04.1.P.Seq | F           | M00039124C:F03 | CH09LNL |
| 798    | 429693  | RTA00002668F.f.05.1.P.Seq | F           | M00032944B:B02 | CH08LNH |
| 799    | 377108  | RTA00002678F.p.04.2.P.Seq | F           | M00039636C:D11 | CH09LNL |
| 800    | 375657  | RTA00002681F.h.13.1.P.Seq | F           | M00039877C:C03 | CH09LNL |
| 801    | 374868  | RTA00002673F.d.08.2.P.Seq | F           | M00039063B:D08 | CH09LNL |
| 802    | 428716  | RTA00002667F.e.08.1.P.Seq | F           | M00032766B:D12 | CH08LNH |
| 803    | 44025   | RTA00002684F.c.01.1.P.Seq | F           | M00040115B:A04 | CH09LNL |
| 804    | 430327  | RTA00002668F.k.11.1.P.Seq | F           | M00033033C:H01 | CH08LNH |
| 805    | 374328  | RTA00002673F.c.24.1.P.Seq | F           | M00039061B:F08 | CH09LNL |
| 806    | 376946  | RTA00002682F.n.10.1.P.Seq | F           | M00040019A:E01 | CH09LNL |
| 807    | 375522  | RTA00002677F.n.08.2.P.Seq | F           | M00039420D:D03 | CH09LNL |
| 808    | 395617  | RTA00002687F.b.15.2.P.Seq | F           | M00039767B:A04 | CH14EDT |
| 809    | 21686   | RTA00002712F.g.05.1.P.Seq | F           | M00026865B:A06 | CH04MAL |
| 810    | 452038  | RTA00002692F.a.09.2.P.Seq | F           | M00042623D:D07 | CH18CON |
| 811    | 25632   | RTA00002711F.g.16.1.P.Seq | F           | M00023042D:D02 | CH03MAH |
| 812    | 152487  | RTA00002663F.e.12.1.P.Seq | F           | M00022181C:D01 | CH03MAH |
| 813    | 378226  | RTA00002680F.g.09.2.P.Seq | F           | M00039797C:G05 | CH09LNL |
| 814    | 402446  | RTA00002686F.c.04.1.P.Seq | F           | M00040133B:B03 | CH13EDT |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 815    | 403642  | RTA00002687F.c.24.2.P.Seq | F            | M00039945C:F09 | CH14EDT   |
| 816    | 186359  | RTA00002713F.g.24.1.P.Seq | F            | M00027379C:B07 | CH04MAL   |
| 817    | 404290  | RTA00002688F.e.04.2.P.Seq | F            | M00040395B:D11 | CH14EDT   |
| 818    | 375443  | RTA00002676F.g.19.2.P.Seq | F            | M00039298B:D03 | CH09LNL   |
| 819    | 380279  | RTA00002673F.i.24.1.P.Seq | F            | M00039082B:A05 | CH09LNL   |
| 820    | 386110  | RTA00002687F.e.06.1.P.Seq | F            | M00039955C:C04 | CH14EDT   |
| 821    | 380279  | RTA00002673F.j.01.1.P.Seq | F            | M00039082B:A05 | CH09LNL   |
| 822    | 386986  | RTA00002675F.p.06.1.P.Seq | F            | M00039266A:B02 | CH09LNL   |
| 823    | 186359  | RTA00002713F.h.01.1.P.Seq | F            | M00027379C:B07 | CH04MAL   |
| 824    | 375611  | RTA00002677F.o.20.2.P.Seq | F            | M00039425D:E12 | CH09LNL   |
| 825    | 378285  | RTA00002679F.h.01.1.P.Seq | F            | M00039681B:H09 | CH09LNL   |
| 826    | 44025   | RTA00002684F.b.24.1.P.Seq | F            | M00040115B:A04 | CH09LNL   |
| 827    | 25240   | RTA00002711F.c.12.1.P.Seq | F            | M00022854A:B03 | CH03MAH   |
| 828    | 403700  | RTA00002687F.g.03.2.P.Seq | F            | M00040207B:D08 | CH14EDT   |
| 829    | 404679  | RTA00002687F.f.07.1.P.Seq | F            | M00040203A:H06 | CH14EDT   |
| 830    | 454806  | RTA00002693F.b.12.2.P.Seq | F            | M00043093C:G11 | CH19COP   |
| 831    | 376829  | RTA00002674F.f.21.2.P.Seq | F            | M00039135D:G02 | CH09LNL   |
| 832    | 456309  | RTA00002694F.d.16.1.P.Seq | F            | M00043518B:D06 | CH20COHLV |
| 833    | 374510  | RTA00002672F.i.17.2.P.Seq | F            | M00039015D:H04 | CH09LNL   |
| 834    | 377232  | RTA00002683F.m.08.2.P.Seq | F            | M00040090B:G09 | CH09LNL   |
| 835    | 375779  | RTA00002672F.j.20.2.P.Seq | F            | M00039025A:H09 | CH09LNL   |
| 836    | 90746   | RTA00002671F.a.07.2.P.Seq | F            | M00033585D:A02 | CH09LNL   |
| 837    | 453002  | RTA00002692F.b.21.2.P.Seq | F            | M00042970C:H10 | CH18CON   |
| 838    | 402863  | RTA00002686F.n.12.1.P.Seq | F            | M00040273B:H12 | CH13EDT   |
| 839    | 402526  | RTA00002686F.p.07.1.P.Seq | F            | M00040286C:C02 | CH13EDT   |
| 840    | 412778  | RTA00002685F.i.07.1.P.Seq | F            | M00039533D:F04 | CH12EDT   |
| 841    | 402273  | RTA00002686F.j.18.1.P.Seq | F            | M00040233C:G05 | CH13EDT   |
| 842    | 374744  | RTA00002670F.i.16.1.P.Seq | F            | M00033427D:F01 | CH09LNL   |
| 843    | 375764  | RTA00002677F.o.18.2.P.Seq | F            | M00039425C:G01 | CH09LNL   |
| 844    | 428218  | RTA00002667F.c.01.1.P.Seq | F            | M00032731C:C07 | CH08LNH   |
| 845    | 374809  | RTA00002675F.h.01.1.P.Seq | F            | M00039230D:D09 | CH09LNL   |
| 846    | 20162   | RTA00002710F.n.20.1.P.Seq | F            | M00022662D:G11 | CH03MAH   |
| 847    | 375782  | RTA00002677F.d.23.2.P.Seq | F            | M00039381C:H08 | CH09LNL   |
| 848    | 372958  | RTA00002672F.c.02.1.P.Seq | F            | M00038639D:F07 | CH09LNL   |
| 849    | 403940  | RTA00002688F.d.07.2.P.Seq | F            | M00040387D:H05 | CH14EDT   |
| 850    | 8490    | RTA00002711F.g.03.1.P.Seq | F            | M00023020C:G08 | CH03MAH   |
| 851    | 374809  | RTA00002675F.g.24.1.P.Seq | F            | M00039230D:D09 | CH09LNL   |
| 852    | 377788  | RTA00002684F.g.24.2.P.Seq | F            | M00040305C:H06 | CH09LNL   |
| 853    | 13847   | RTA00002711F.f.09.1.P.Seq | F            | M00022976C:F04 | CH03MAH   |
| 854    | 374172  | RTA00002673F.k.16.1.P.Seq | F            | M00039097D:D06 | CH09LNL   |
| 855    | 380314  | RTA00002682F.l.07.1.P.Seq | F            | M00040009D:B07 | CH09LNL   |
| 856    | 47231   | RTA00002714F.b.15.1.P.Seq | F            | M00027813C:F01 | CH04MAL   |
| 857    | 400287  | RTA00002685F.k.10.1.P.Seq | F            | M00039584C:C01 | CH12EDT   |
| 858    | 400533  | RTA00002685F.a.02.2.P.Seq | F            | M00039181D:E05 | CH12EDT   |
| 859    | 447594  | RTA00002689F.c.07.1.P.Seq | F            | M00042696B:E05 | CH15CON   |
| 860    | 147357  | RTA00002711F.e.15.1.P.Seq | F            | M00022928B:C01 | CH03MAH   |
| 861    | 401141  | RTA00002685F.o.22.2.P.Seq | F            | M00039642D:B12 | CH12EDT   |
| 862    | 404620  | RTA00002687F.c.03.1.P.Seq | F            | M00039770A:G11 | CH14EDT   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY   |
|--------|---------|---------------------------|-------------|-----------------|-----------|
| 863    | 24360   | RTA00002709F.l.20.1.P.Seq | F           | M00007149A:G02  | CH02COH   |
| 864    | 380618  | RTA00002673F.j.12.1.P.Seq | F           | M00039084C:G07  | CH09LNL   |
| 865    | 448446  | RTA00002690F.d.09.3.P.Seq | F           | M00042797D:D10  | CH16COP   |
| 866    | 402313  | RTA00002686F.f.18.1.P.Seq | F           | M00040174D:G03  | CH13EDT   |
| 867    | 273151  | RTA00002685F.c.05.2.P.Seq | F           | M00039374C:H02  | CH12EDT   |
| 868    | 404172  | RTA00002687F.d.17.2.P.Seq | F           | M00039951B:B12  | CH14EDT   |
| 869    | 263630  | RTA00002694F.e.10.1.P.Seq | F           | M00043637C:H01  | CH20COHLV |
| 870    | 404277  | RTA00002687F.d.18.1.P.Seq | F           | M00039951B:C03  | CH14EDT   |
| 871    | 403557  | RTA00002687F.d.10.1.P.Seq | F           | M00039948A:E03  | CH14EDT   |
| 872    | 375161  | RTA00002676F.m.24.2.P.Seq | F           | M00039319B:H12  | CH09LNL   |
| 873    | 376829  | RTA00002674F.f.21.1.P.Seq | F           | M00039135D:G02  | CH09LNL   |
| 874    | 372958  | RTA00002672F.c.02.2.P.Seq | F           | M00038639D:F07  | CH9LNL    |
| 875    | 21578   | RTA00002709F.a.24.1.P.Seq | F           | M00005351C:G05  | CH02COH   |
| 876    | 402506  | RTA00002686F.b.17.1.P.Seq | F           | M00039760B:B08  | CH13EDT   |
| 877    | 141731  | RTA00002713F.b.04.1.P.Seq | F           | M00027212D:E03  | CH04MAL   |
| 878    | 37411   | RTA00002661F.e.11.1.P.Seq | F           | M00003770A:E05  | CH01COH   |
| 879    | 372537  | RTA00002670F.c.05.2.P.Seq | F           | M00033345D:A09  | CH09LNL   |
| 880    | 380834  | RTA00002670F.c.08.2.P.Seq | F           | M00033346C:A05  | CH09LNL   |
| 881    | 401492  | RTA00002685F.n.17.2.P.Seq | F           | M00039609D:F07  | CH12EDT   |
| 882    | 99998   | RTA00002662F.b.23.2.P.Seq | F           | M00006712C:H09  | CH02COH   |
| 883    | 404311  | RTA00002688F.d.21.2.P.Seq | F           | M00040394A:D04  | CH14EDT   |
| 884    | 231084  | RTA00002664F.c.18.2.P.Seq | F           | M00026918B:D01  | CH04MAL   |
| 885    | 447679  | RTA00002689F.b.11.3.P.Seq | F           | M00042560A:F12  | CH15CON   |
| 886    | 377012  | RTA00002682F.d.17.1.P.Seq | F           | M00039936C:C05  | CH09LNL   |
| 887    | 226207  | RTA00002664F.d.21.2.P.Seq | F           | M00027035D:C06  | CH04MAL   |
| 888    | 446183  | RTA00002689F.a.12.1.P.Seq | F           | M00042634A:A05  | CH15CON   |
| 889    | 428508  | RTA00002666F.c.24.1.P.Seq | F           | M00032545B:H09  | CH08LNH   |
| 890    | 157648  | RTA00002714F.b.20.1.P.Seq | F           | M00027818C:C07  | CH04MAL   |
| 891    | 404609  | RTA00002688F.b.15.2.P.Seq | F           | M00040377C:G07  | CH14EDT   |
| 892    | 400464  | RTA00002685F.l.10.1.P.Seq | F           | M00039590D:D02  | CH12EDT   |
| 893    | 379108  | RTA00002685F.l.12.1.P.Seq | F           | M00039591C:D06  | CH12EDT   |
| 894    | 374639  | RTA00002676F.d.21.2.P.Seq | F           | M00039284D:B12  | CH09LNL   |
| 895    | 380674  | RTA00002673F.j.14.2.P.Seq | F           | M00039084C:H04  | CH09LNL   |
| 896    | 380674  | RTA00002673F.j.14.1.P.Seq | F           | M000F39084C:H04 | CH09LNL   |
| 897    | 188972  | RTA00002664F.d.20.2.P.Seq | F           | M00027030C:H06  | CH04MAL   |
| 898    | 402835  | RTA00002686F.c.01.1.P.Seq | F           | M00040131D:G08  | CH13EDT   |
| 899    | 403774  | RTA00002687F.d.08.2.P.Seq | F           | M00039947C:G03  | CH14EDT   |
| 900    | 374606  | RTA00002673F.j.23.2.P.Seq | F           | M00039096A:A05  | CH09LNL   |
| 901    | 192535  | RTA00002663F.m.14.1.P.Seq | F           | M00022925C:A08  | CH03MAH   |
| 902    | 377926  | RTA00002680F.l.16.2.P.Seq | F           | M00039820B:B06  | CH09LNL   |
| 903    | 186055  | RTA00002712F.i.11.1.P.Seq | F           | M00026926A:E10  | CH04MAL   |
| 904    | 380498  | RTA00002684F.f.11.2.P.Seq | F           | M00040129D:E10  | CH09LNL   |
| 905    | 400236  | RTA00002685F.i.18.2.P.Seq | F           | M00039561A:B07  | CH12EDT   |
| 906    | 401070  | RTA00002688F.d.12.2.P.Seq | F           | M00040390A:H02  | CH14EDT   |
| 907    | 452622  | RTA00002692F.b.14.2.P.Seq | F           | M00042962D:C05  | CH18CON   |
| 908    | 235052  | RTA00002692F.a.15.1.P.Seq | F           | M00042626B:D08  | CH18CON   |
| 909    | 452221  | RTA00002692F.c.13.2.P.Seq | F           | M00042986C:G12  | CH18CON   |
| 910    | 404581  | RTA00002687F.g.11.2.P.Seq | F           | M00040208D:G09  | CH14EDT   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 911    | 376925  | RTA00002687F.e.14.2.P.Seq | F            | M00039957C:C09 | CH14EDT   |
| 912    | 400287  | RTA00002685F.k.10.2.P.Seq | F            | M00039584C:C01 | CH12EDT   |
| 913    | 403242  | RTA00002687F.l.05.2.P.Seq | F            | M00040323B:C12 | CH14EDT   |
| 914    | 453313  | RTA00002693F.a.07.2.P.Seq | F            | M00042614B:B05 | CH19COP   |
| 915    | 452633  | RTA00002692F.f.11.2.P.Seq | F            | M00043067D:D10 | CH18CON   |
| 916    | 447679  | RTA00002689F.b.11.1.P.Seq | F            | M00042560A:F12 | CH15CON   |
| 917    | 452398  | RTA00002692F.f.17.1.P.Seq | F            | M00043125C:A11 | CH18CON   |
| 918    | 449797  | RTA00002691F.b.22.3.P.Seq | F            | M00043334B:A10 | CH17COHLV |
| 919    | 403916  | RTA00002687F.j.11.2.P.Seq | F            | M00040314D:H05 | CH14EDT   |
| 920    | 236906  | RTA00002693F.d.05.2.P.Seq | F            | M00043154A:B07 | CH19COP   |
| 921    | 40416   | RTA00002687F.e.20.2.P.Seq | F            | M00039958C:B09 | CH14EDT   |
| 922    | 386110  | RTA00002687F.e.06.2.P.Seq | F            | M00039955C:C04 | CH14EDT   |
| 923    | 451512  | RTA00002691F.b.02.3.P.Seq | F            | M00043305B:G02 | CH17COHLV |
| 924    | 400517  | RTA00002687F.k.15.2.P.Seq | F            | M00040320D:F02 | CH14EDT   |
| 925    | 403578  | RTA00002687F.i.01.2.P.Seq | F            | M00040296D:E09 | CH14EDT   |
| 926    | 403578  | RTA00002687F.h.24.2.P.Seq | F            | M00040296D:E09 | CH14EDT   |
| 927    | 403371  | RTA00002687F.h.19.2.P.Seq | F            | M00040294D:D12 | CH14EDT   |
| 928    | 452531  | RTA00002692F.f.16.1.P.Seq | F            | M00043125A:B11 | CH18CON   |
| 929    | 454453  | RTA00002693F.f.15.2.P.Seq | F            | M00043215A:D02 | CH19COP   |
| 930    | 238270  | RTA00002692F.e.07.2.P.Seq | F            | M00043028A:G05 | CH18CON   |
| 931    | 14583   | RTA00002687F.f.08.2.P.Seq | F            | M00040203B:A05 | CH14EDT   |
| 932    | 400464  | RTA00002685F.l.10.2.P.Seq | F            | M00039590D:D02 | CH12EDT   |
| 933    | 404642  | RTA00002687F.f.02.2.P.Seq | F            | M00040201C:G11 | CH14EDT   |
| 934    | 380413  | RTA00002680F.k.19.1.P.Seq | F            | M00039816C:D05 | CH09LNL   |
| 935    | 287963  | RTA00002693F.c.20.2.P.Seq | F            | M00043148C:A09 | CH19COP   |
| 936    | 20847   | RTA00002710F.d.09.1.P.Seq | F            | M00021852D:A05 | CH03MAH   |
| 937    | 456531  | RTA00002694F.b.18.1.P.Seq | F            | M00043446C:E12 | CH20COHLV |
| 938    | 450463  | RTA00002694F.a.12.1.P.Seq | F            | M00042596C:D07 | CH20COHLV |
| 939    | 456713  | RTA00002694F.d.13.1.P.Seq | F            | M00043513D:G08 | CH20COHLV |
| 940    | 455508  | RTA00002694F.a.15.1.P.Seq | F            | M00042597B:E12 | CH20COHLV |

TABLE 2

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID        | LIBRARY  |
|--------|---------|---------------------------|--------------|-----------------|----------|
| 1      | 10600   | RTA00002891F.i.07.1.P.Seq | F            | M00003758B:D07  | CH01COH  |
| 2      | 18827   | RTA00002900F.p.12.1.P.Seq | F            | M00005413D:A05  | CH02COH  |
| 3      | 1759    | RTA00002923F.f.23.1.P.Seq | F            | M000039248C:A08 | CH09LNL  |
| 4      | 10924   | RTA00002907F.k.12.1.P.Seq | F            | M00022224A:C07  | CH03MAH  |
| 5      | 45331   | RTA00002903F.l.10.1.P.Seq | F            | M00007037D:D10  | CH02COH  |
| 6      | 42233   | RTA00002912F.g.24.1.P.Seq | F            | M00027359B:A06  | CH04MAL  |
| 7      | 7211    | RTA00002909F.h.06.1.P.Seq | F            | M00022634A:C07  | CH03MAH  |
| 8      | 21395   | RTA00002890F.k.16.1.P.Seq | F            | M00001637D:C12  | CH01COH  |
| 9      | 3093    | RTA0002923F.e.03.1.P.Seq  | F            | M00039225A:D11  | CH09LNL  |
| 10     | 15806   | RTA00002894F.f.07.1.P.Seq | F            | M00003991A:C11  | CH01COH  |
| 11     | 19739   | RTA00002896F.d.12.1.P.Seq | F            | M00004147C:E01  | CH01COH  |
| 12     | 140879  | RTA00002905F.c.17.1.P.Seq | F            | M00007985C:D08  | CH03MAH  |
| 13     | 29706   | RTA00002908F.l.22.1.P.Seq | F            | M00022487B:A08  | CH03MAH  |
| 14     | 109581  | RTA00002918F.i.08.1.P.Seq | F            | M00032908A:D08  | CH08LNL  |
| 15     | 25009   | RTA00002906F.k.11.1.P.Seq | F            | M00022016B:F01  | CH03MAH  |
| 16     | 8328    | RTA0002888F.e.07.1.P.Seq  | F            | M00001451C:E10  | CH01COH  |
| 17     | 15045   | RTA00002887F.e.16.1.P.Seq | F            | M00001393C:E08  | CH01COH  |
| 18     | 21216   | RTA00002898F.p.22.1.P.Seq | F            | M00004416B:G10  | CH01COH  |
| 19     | 185754  | RTA00002912F.l.09.1.P.Seq | F            | M00027506B:G01  | CH04MAL  |
| 20     | 11881   | RTA00002909F.h.10.1.P.Seq | F            | M00022638A:D03  | CH03MAH  |
| 21     | 185989  | RTA00002910F.h.12.1.P.Seq | F            | M00022924C:F04  | CH03MAH  |
| 22     | 9667    | RTA00002923F.a.03.1.P.Seq | F            | M00039162D:C04  | CH09LNL  |
| 23     | 15817   | RTA00002903F.o.03.1.P.Seq | F            | M00007103D:C02  | CH02COH  |
| 24     | 10198   | RTA00002923F.j.09.1.P.Seq | F            | M00039294C:B09  | CH09LNL  |
| 25     | 6355    | RTA00002894F.p.12.1.P.Seq | F            | M00004055D:D05  | CH01COH  |
| 26     | 12227   | RTA00002909F.e.18.1.P.Seq | F            | M00022601B:G06  | CH03MAH  |
| 27     | 11047   | RTA00002893F.o.06.1.P.Seq | F            | M00003960D:C12  | CH01COH  |
| 28     | 1870    | RTA00002910F.m.08.1.P.Seq | F            | M00023020C:H03  | CH03MAH  |
| 29     | 20065   | RTA00002908F.m.09.1.P.Seq | F            | M00022491A:A08  | CH03MAH  |
| 30     | 19454   | RTA00002900F.m.23.1.P.Seq | F            | M00005379A:D10  | CH02COH  |
| 31     | 48048   | RTA00002922F.m.13.1.P.Seq | F            | M00039:24D:H01  | CH09LNL  |
| 32     | 19799   | RTA00002908F.h.19.1.P.Seq | F            | M00022449D:F08  | CH03MAH  |
| 33     | 185562  | RTA00002911F.m.07.1.P.Seq | F            | M00027093A:H02  | CH04MAL  |
| 34     | 24214   | RTA00002891F.k.19.1.P.Seq | F            | M00003764D:F07  | CH01COH  |
| 35     | 5172    | RTA00002908F.p.22.1.P.Seq | F            | M00022525B:D09  | CH03MAH  |
| 36     | 50495   | RTA00002898F.c.16.1.P.Seq | F            | M00004321C:C11  | CH01COH  |
| 37     | 43287   | RTA00002908F.k.16.1.P.Seq | F            | M00022470D:B02  | CH03MAH  |
| 38     | 15324   | RTA00002905F.p.20.1.P.Seq | F            | M00021697C:B07  | CH03MAH  |
| 39     | 22157   | RTA00002888F.g.07.1.P.Seq | F            | M00001461D:B10  | CH01COH  |
| 40     | 15249   | RTA00002915F.l.08.1.P.Seq | F            | M00032489B:G12  | CH08LNL  |
| 41     | 2764    | RTA00002925F.c.11.1.P.Seq | F            | M00039829B:E01  | CHK09LNL |
| 42     | 23838   | RTA00002889F.b.14.1.P.Seq | F            | M00001548B:D10  | CH01COH  |
| 43     | 11074   | RTA00002899F.g.22.1.P.Seq | F            | M00004603C:C10  | CH01COH  |
| 44     | 18367   | RTA00002922F.b.09.1.P.Seq | F            | M00038619D:C12  | CH09LNL  |
| 45     | 21703   | RTA00002903F.m.08.1.P.Seq | F            | M00007059B:D07  | CH02COH  |
| 46     | 21470   | RTA00002895F.c.14.1.P.Seq | F            | M00004067B:D03  | CH01COH  |
| 47     | 15492   | RTA00002907F.p.06.1.P.Seq | F            | M00022282B:C09  | CH03MAH  |
| 48     | 4022    | RTA00002897F.i.22.1.P.Seq | F            | M00004269B:B04  | CH01COH  |
| 49     | 21579   | RTA00002891F.e.03.1.P.Seq | F            | M00001686B:H01  | CH01COH  |
| 50     | 186283  | RTA00002913F.c.06.1.P.Seq | F            | M00027801B:D07  | CH04MAL  |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|--------------|----------------|---------|
| 51     | 5410    | RTA00002921F.i.08.1.P.Seq | F            | M00033445D:G03 | CH09LNL |
| 52     | 22420   | RTA00002901F.e.19.1.P.Seq | F            | M00005474C:H09 | CH02COH |
| 53     | 140553  | RTA00002916F.n.02.1.P.Seq | F            | M00032638B:F02 | CH08LNH |
| 54     | 23849   | RTA00002887F.a.22.1.P.Seq | F            | M00001386B:F11 | CH01COH |
| 55     | 21945   | RTA00002895F.l.22.1.P.Seq | F            | M00004103C:E10 | CH01COH |
| 56     | 7867    | RTA00002901F.p.08.1.P.Seq | F            | M00005710B:H03 | CH02COH |
| 57     | 14533   | RTA00002896F.l.01.1.P.Seq | F            | M00004179C:B06 | CH01COH |
| 58     | 5790    | RTA00002919F.g.17.1.P.Seq | F            | M00033080C:A07 | CH08LNH |
| 59     | 186153  | RTA00002911F.i.24.1.P.Seq | F            | M00027017A:B09 | CH04MAL |
| 60     | 10561   | RTA00002899F.h.08.1.P.Seq | F            | M00004606D:H09 | CH01COH |
| 61     | 24572   | RTA00002893F.l.08.1.P.Seq | F            | M00003926A:F11 | CH01COH |
| 62     | 13138   | RTA00002888F.m.03.1.P.Seq | F            | M00001488C:A03 | CH01COH |
| 63     | 6701    | RTA00002922F.g.18.1.P.Seq | F            | M00039055C:A01 | CH09LNL |
| 64     | 12751   | RTA00002904F.c.10.1.P.Seq | F            | M00007202B:F01 | CH02COH |
| 65     | 3583    | RTA00002916F.n.21.1.P.Seq | F            | M00032644C:B05 | CH08LNH |
| 66     | 12673   | RTA00002901F.d.24.1.P.Seq | F            | M00005463A:G02 | CH02COH |
| 67     | 15243   | RTA00002901F.l.21.1.P.Seq | F            | M00005623B:G01 | CH02COH |
| 68     | 21022   | RTA00002922F.k.24.1.P.Seq | F            | M00039111A:C12 | CH09LNL |
| 69     | 36596   | RTA00002919F.g.24.1.P.Seq | F            | M00033081D:D11 | CH08LNH |
| 70     | 4932    | RTA00002890F.c.14.1.P.Seq | F            | M00001596A:D02 | CH01COH |
| 71     | 42413   | RTA00002900F.o.14.1.P.Seq | F            | M00005401D:F09 | CH02COH |
| 72     | 1090    | RTA00002918F.g.20.1.P.Seq | F            | M00032892C:C12 | CH08LNH |
| 73     | 44737   | RTA00002901F.a.20.1.P.Seq | F            | M00005434A:C03 | CH02COH |
| 74     | 4183    | RTA00002918F.n.23.1.P.Seq | F            | M00032988B:G01 | CH08LNH |
| 75     | 41882   | RTA00002902F.d.12.1.P.Seq | F            | M00006586D:D04 | CH02COH |
| 76     | 500     | RTA00002925F.o.18.1.P.Seq | F            | M00040034A:E06 | CH09LNL |
| 77     | 5435    | RTA00002921F.f.20.1.P.Seq | F            | M00033420B:E08 | CH09LNL |
| 78     | 15829   | RTA00002900F.i.01.1.P.Seq | F            | M00005314A:G10 | CH02COH |
| 79     | 154083  | RTA00002907F.a.06.1.P.Seq | F            | M00022096D:A03 | CH03MAH |
| 80     | 24381   | RTA00002910F.i.16.1.P.Seq | F            | M00022953B:D06 | CH03MAH |
| 81     | 107940  | RTA00002930F.f.07.1.P.Seq | F            | M00055735A:H08 | CH15CON |
| 82     | 24761   | RTA00002902F.l.21.1.P.Seq | F            | M00006756C:A02 | CH02COH |
| 83     | 10734   | RTA00002924F.e.02.1.P.Seq | F            | M00039457D:C02 | CH09LNL |
| 84     | 40540   | RTA00002897F.p.23.1.P.Seq | F            | M0004303C:C05  | CH01COH |
| 85     | 23692   | RTA00002930F.i.07.1.P.Seq | F            | M00056057C:F06 | CH15CON |
| 86     | 7896    | RTA00002906F.j.08.1.P.Seq | F            | M00021998B:D09 | CH03MAH |
| 87     | 24387   | RTA00002896F.e.09.1.P.Seq | F            | M00004151B:A07 | CH01COH |
| 88     | 2420    | RTA00002889F.h.02.1.P.Seq | F            | M00001546B:C11 | CH01COH |
| 89     | 10431   | RTA00002887F.p.07.1.P.Seq | F            | M00001429B:G05 | CH01COH |
| 90     | 14665   | RTA00002908F.g.07.1.P.Seq | F            | M00022425A:C09 | CH03MAH |
| 91     | 10302   | RTA00002906F.o.03.1.P.Seq | F            | M00022081A:B07 | CH03MAH |
| 92     | 28436   | RTA00002908F.f.08.1.P.Seq | F            | M00022415C:D12 | CH03MAH |
| 93     | 17829   | RTA00002889F.g.11.1.P.Seq | F            | M00001544B:E06 | CH01COH |
| 94     | 10390   | RTA00002906F.e.13.1.P.Seq | F            | M00021923A:B12 | CH03MAH |
| 95     | 11619   | RTA00002913F.c.07.1.P.Seq | F            | M00027806C:H05 | CH04MAL |
| 96     | 6890    | RTA00002918F.m.19.1.P.Seq | F            | M00032979D:H07 | CH08LNH |
| 97     | 10110   | RTA00002897F.g.13.1.P.Seq | F            | M00004245C:G10 | CH01COH |
| 98     | 21511   | RTA00002892F.h.24.2.P.Seq | F            | M00003821C:E12 | CH01COH |
| 99     | 9287    | RTA00002899F.h.14.1.P.Seq | F            | M00004608A:C10 | CH01COH |
| 100    | 16575   | RTA00002895F.n.02.1.P.Seq | F            | M00004110D:F09 | CH01COH |
| 101    | 16857   | RTA00002899F.h.17.1.P.Seq | F            | M00004609A:E09 | CH01COH |
| 102    | 33329   | RTA00002891F.p.14.1.P.Seq | F            | M00003787D:A10 | CH01COH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|-------------|----------------|---------|
| 103    | 40652   | RTA00002896F.f.21.1.P.Seq | F           | M00004158B:E03 | CH01COH |
| 104    | 8070    | RTA00002925F.c.09.1.P.Seq | F           | M00039828B:H06 | CH09LNL |
| 105    | 15880   | RTA00002887F.g.11.1.P.Seq | F           | M00001397C:H08 | CH01COH |
| 106    | 87418   | RTA00002901F.g.10.1.P.Seq | F           | M00005500A:D04 | CH02COH |
| 107    | 9961    | RTA00002903F.l.21.1.P.Seq | F           | M00007046D:C09 | CH02COH |
| 108    | 9966    | RTA00002911F.o.20.1.P.Seq | F           | M00027168B:H08 | CH04MAL |
| 109    | 17513   | RTA00002906F.d.03.1.P.Seq | F           | M00021896D:A05 | CH03MAH |
| 110    | 24835   | RTA00002924F.f.18.1.P.Seq | F           | M00039554D:B09 | CH09LNL |
| 111    | 15200   | RTA00002891F.i.08.1.P.Seq | F           | M00003758B:F06 | CH01COH |
| 112    | 124098  | RTA00002905F.e.23.1.P.Seq | F           | M00008020D:D05 | CH03MAH |
| 113    | 3786    | RTA00002901F.e.05.1.P.Seq | F           | M00005466C:B01 | CH02COH |
| 114    | 154121  | RTA00002906F.p.02.1.P.Seq | F           | M00022088B:F10 | CH03MAH |
| 115    | 5746    | RTA00002918F.l.09.1.P.Seq | F           | M00032945D:B07 | CH08LNL |
| 116    | 33700   | RTA00002901F.e.09.1.P.Seq | F           | M00005468A:C04 | CH02COH |
| 117    | 5660    | RTA00002890F.b.11.1.P.Seq | F           | M00001591B:H05 | CH01COH |
| 118    | 22732   | RTA00002924F.o.23.1.P.Seq | F           | M00039785C:H12 | CH09LNL |
| 119    | 14720   | RTA00002892F.j.05.1.P.Seq | F           | M00003825A:H10 | CH01COH |
| 120    | 13658   | RTA00002896F.c.04.1.P.Seq | F           | M00004143B:B04 | CH01COH |
| 121    | 23150   | RTA00002887F.a.01.1.P.Seq | F           | M00001384A:A07 | CH01COH |
| 122    | 11970   | RTA00002903F.k.03.1.P.Seq | F           | M00007002C:A10 | CH02COH |
| 123    | 10686   | RTA00002915F.p.02.2.P.Seq | F           | M00032519D:F08 | CH08LNL |
| 124    | 9588    | RTA00002923F.m.10.1.P.Seq | F           | M00039331B:F09 | CH09LNL |
| 125    | 8500    | RTA00002925F.e.21.1.P.Seq | F           | M00039860D:B02 | CH09LNL |
| 126    | 8615    | RTA00002907F.l.17.1.P.Seq | F           | M00022238C:G04 | CH03MAH |
| 127    | 7524    | RTA00002886F.e.16.1.P.Seq | F           | M00001348B:B03 | CH01COH |
| 128    | 325     | RTA00002912F.g.02.1.P.Seq | F           | M00027332B:H09 | CH04MAL |
| 129    | 10214   | RTA00002889F.o.01.1.P.Seq | F           | M00001570A:B07 | CH01COH |
| 130    | 23534   | RTA00002889F.c.22.1.P.Seq | F           | M00001533D:A01 | CH01COH |
| 131    | 7473    | RTA00002893F.o.20.1.P.Seq | F           | M00003965D:D11 | CH01COH |
| 132    | 185625  | RTA00002912F.f.10.1.P.Seq | F           | M00027314D:E02 | CH04MAL |
| 133    | 3920    | RTA00002917F.m.07.1.P.Seq | F           | M00032773D:F08 | CH08LNL |
| 134    | 8458    | RTA00002889F.m.02.1.P.Seq | F           | M00001562D:B07 | CH01COH |
| 135    | 20263   | RTA00002906F.n.08.1.P.Seq | F           | M00022073C:C07 | CH03MAH |
| 136    | 186141  | RTA00002912F.f.04.1.P.Seq | F           | M00027311A:H09 | CH04MAL |
| 137    | 4852    | RTA00002919F.i.15.1.P.Seq | F           | M00033150B:E02 | CH08LNL |
| 138    | 2146    | RTA00002926F.a.21.1.P.Seq | F           | M00040061C:C08 | CH09LNL |
| 139    | 3522    | RTA00002897F.h.23.1.P.Seq | F           | M00004263C:D03 | CH01COH |
| 140    | 20027   | RTA00002909F.g.21.1.P.Seq | F           | M00022627B:H03 | CH03MAH |
| 141    | 15650   | RTA00002925F.g.04.1.P.Seq | F           | M00039874A:B06 | CH09LNL |
| 142    | 21031   | RTA00002901F.h.21.1.P.Seq | F           | M00005520B:H05 | CH02COH |
| 143    | 95610   | RTA00002909F.h.18.1.P.Seq | F           | M00022642A:G08 | CH03MAH |
| 144    | 903     | RTA00002912F.o.03.1.P.Seq | F           | M00027591A:E04 | CH04MAL |
| 145    | 17284   | RTA00002916F.k.18.1.P.Seq | F           | M00032620B:F06 | CH08LNL |
| 146    | 15556   | RTA00002895F.m.01.1.P.Seq | F           | M00004104A:A12 | CH01COH |
| 147    | 11013   | RTA00002897F.b.08.1.P.Seq | F           | M00004214D:A05 | CH01COH |
| 148    | 15358   | RTA00002903F.n.18.1.P.Seq | F           | M00007098A:E10 | CH02COH |
| 149    | 10792   | RTA00002894F.a.10.1.P.Seq | F           | M00003974C:E11 | CH01COH |
| 150    | 25507   | RTA00002901F.n.07.1.P.Seq | F           | M00005643D:A05 | CH02COH |
| 151    | 19978   | RTA00002908F.k.21.1.P.Seq | F           | M00022472D:B01 | CH03MAH |
| 152    | 31133   | RTA00002905F.f.23.1.P.Seq | F           | M00008045C:A05 | CH03MAH |
| 153    | 4503    | RTA00002893F.h.14.1.P.Seq | F           | M00003906A:C02 | CH01COH |
| 154    | 9017    | RTA00002886F.i.03.1.P.Seq | F           | M00001358A:E08 | CH01COH |



| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|-------------|-----------------|---------|
| 155    | 6635    | RTA00002895F.a.07.1.P.Seq | F           | M00004057D:G01  | CH01COH |
| 156    | 10220   | RTA00002921F.d.09.1.P.Seq | F           | M00033360C:A03  | CH09LNL |
| 157    | 9831    | RTA00002896F.b.08.1.P.Seq | F           | M00004139B:F01  | CH01COH |
| 158    | 17291   | RTA00002888F.c.10.1.P.Seq | F           | M00001442A:F08  | CH01COH |
| 159    | 9802    | RTA00002916F.1.12.1.P.Seq | F           | M00032628C:B06  | CH08LNL |
| 160    | 133828  | RTA00002908F.b.21.1.P.Seq | F           | M00022374C:E11  | CH03MAH |
| 161    | 25870   | RTA00002909F.m.23.1.P.Seq | F           | M00022702D:E02  | CH03MAH |
| 162    | 27324   | RTA00002912F.i.22.1.P.Seq | F           | M00027433B:D12  | CH04MAL |
| 163    | 98159   | RTA00002910F.f.19.1.P.Seq | F           | M00022897B:F06  | CH03MAH |
| 164    | 21264   | RTA00002903F.d.15.1.P.Seq | F           | M00006904D:A02  | CH02COH |
| 165    | 186199  | RTA00002911F.b.13.2.P.Seq | F           | M00023394D:D10  | CH04MAL |
| 166    | 28794   | RTA00002887F.i.14.1.P.Seq | F           | M00001403C:B03  | CH01COH |
| 167    | 23180   | RTA00002895F.i.17.1.P.Seq | F           | M00004093A:C03  | CH01COH |
| 168    | 21022   | RTA00002922F.1.01.1.P.Seq | F           | M000039111A:C12 | CH09LNL |
| 169    | 14370   | RTA00002893F.i.17.1.P.Seq | F           | M00003911C:A09  | CH01COH |
| 170    | 4804    | RTA00002918F.a.19.1.P.Seq | F           | M00032826C:D10  | CH08LNL |
| 171    | 7066    | RTA00002919F.o.07.1.P.Seq | F           | M00033246A:H12  | CH08LNL |
| 172    | 48227   | RTA00002903F.o.18.1.P.Seq | F           | M00007117A:C11  | CH02COH |
| 173    | 20171   | RTA00002886F.i.15.1.P.Seq | F           | M00001359A:H10  | CH01COH |
| 174    | 10555   | RTA00002894F.p.10.1.P.Seq | F           | M00004055C:B10  | CH01COH |
| 175    | 12523   | RTA00002914F.m.08.1.P.Seq | F           | M00028361B:H08  | CH08LNL |
| 176    | 23767   | RTA00002896F.i.21.1.P.Seq | F           | M00004171B:B03  | CH01COH |
| 177    | 16849   | RTA00002918F.b.07.1.P.Seq | F           | M00032829D:A05  | CH08LNL |
| 178    | 185866  | RTA00002911F.c.18.2.P.Seq | F           | M00026818C:E01  | CH04MAL |
| 179    | 29927   | RTA00002899F.b.20.1.P.Seq | F           | M00004443C:F07  | CH01COH |
| 180    | 21975   | RTA00002902F.a.01.1.P.Seq | F           | M00005743D:A12  | CH02COH |
| 181    | 24456   | RTA00002903F.b.20.1.P.Seq | F           | M00006877C:F11  | CH02COH |
| 182    | 6034    | RTA00002901F.a.12.1.P.Seq | F           | M00005423A:C11  | CH02COH |
| 183    | 11362   | RTA00002887F.h.06.1.P.Seq | F           | M00001399C:A01  | CH01COH |
| 184    | 20671   | RTA00002905F.a.22.1.P.Seq | F           | M00007947A:B06  | CH03MAH |
| 185    | 8059    | RTA00002917F.b.02.1.P.Seq | F           | M000032671B:D06 | CH08LNL |
| 186    | 12037   | RTA00002897F.d.11.1.P.Seq | F           | M00004229B:B06  | CH01COH |
| 187    | 13209   | RTA00002897F.d.20.1.P.Seq | F           | M00004230D:B05  | CH01COH |
| 188    | 23660   | RTA00002915F.i.21.1.P.Seq | F           | M00031416D:H05  | CH08LNL |
| 189    | 4747    | RTA00002919F.c.23.1.P.Seq | F           | M00033041A:B11  | CH08LNL |
| 190    | 24532   | RTA00002919F.m.16.1.P.Seq | F           | M00033218C:F07  | CH08LNL |
| 191    | 8576    | RTA00002890F.h.13.1.P.Seq | F           | M00001616D:F03  | CH01COH |
| 192    | 12056   | RTA00002893F.g.12.1.P.Seq | F           | M00003900C:D12  | CH01COH |
| 193    | 895     | RTA00002921F.b.11.1.P.Seq | F           | M00033303C:F09  | CH09LNL |
| 194    | 7212    | RTA00002897F.j.04.1.P.Seq | F           | M00004270A:E09  | CH01COH |
| 195    | 108296  | RTA00002907F.h.20.1.P.Seq | F           | M00022193C:C09  | CH03MAH |
| 196    | 115713  | RTA00002906F.a.22.1.P.Seq | F           | M00021852C:H02  | CH03MAH |
| 197    | 7334    | RTA00002910F.1.08.1.P.Seq | F           | M00023004C:A01  | CH03MAH |
| 198    | 1090    | RTA00002918F.g.20.2.P.Seq | F           | M00032892C:C12  | CH08LNL |
| 199    | 7913    | RTA00002886F.i.13.1.P.Seq | F           | M00001362A:F09  | CH01COH |
| 200    | 12139   | RTA00002923F.o.02.1.P.Seq | F           | M00039349D:B11  | CH09LNL |
| 201    | 17148   | RTA00002895F.1.14.1.P.Seq | F           | M00004102A:E03  | CH01COH |
| 202    | 2314    | RTA00002891F.g.10.1.P.Seq | F           | M00001771B:E06  | CH01COH |
| 203    | 18660   | RTA00002903F.m.03.1.P.Seq | F           | M00007048C:A12  | CH02COH |
| 204    | 21799   | RTA00002900F.n.09.1.P.Seq | F           | M00005385D:F07  | CH02COH |
| 205    | 16612   | RTA00002893F.i.17.1.P.Seq | F           | M00003915C:D10  | CH01COH |
| 206    | 168067  | RTA00002909F.a.17.1.P.Seq | F           | M00022537B:C06  | CH03MAH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|-------------|-----------------|---------|
| 207    | 21197   | RTA00002891F.c.19.1.P.Seq | F           | M00001680A:A01  | CH01COH |
| 208    | 45015   | RTA00002905F.o.17.1.P.Seq | F           | M00021678D:H04  | CH03MAH |
| 209    | 15381   | RTA00002903F.m.19.1.P.Seq | F           | M00007070C:C01  | CH02COH |
| 210    | 1402    | RTA00002896F.i.04.1.P.Seq | F           | M00004166C:B10  | CH01COH |
| 211    | 69026   | RTA00002893F.f.05.1.P.Seq | F           | M00003887C:E09  | CH01COH |
| 212    | 33119   | RTA00002922F.i.07.1.P.Seq | F           | M00039067A:C05  | CH09LNL |
| 213    | 1166    | RTA00002891F.c.05.1.P.Seq | F           | M00001677B:H08  | CH01COH |
| 214    | 14345   | RTA00002891F.b.06.1.P.Seq | F           | M00001671C:F03  | CH01COH |
| 215    | 10589   | RTA00002908F.d.16.1.P.Seq | F           | M00022392B:F01  | CH03MAH |
| 216    | 13281   | RTA00002916F.p.01.1.P.Seq | F           | M00032652C:C07  | CH08LNL |
| 217    | 13281   | RTA00002916F.o.24.1.P.Seq | F           | M00032652C:C07  | CH08LNL |
| 218    | 12248   | RTA00002898F.h.08.1.P.Seq | F           | M00004351B:G07  | CH01COH |
| 219    | 164955  | RTA00002909F.n.09.1.P.Seq | F           | M00022706D:G08  | CH03MAH |
| 220    | 15686   | RTA00002907F.f.24.1.P.Seq | F           | M00022170C:C01  | CH03MAH |
| 221    | 15686   | RTA00002907F.g.01.1.P.Seq | F           | M00022170C:C01  | CH03MAH |
| 222    | 34228   | RTA00002891F.d.14.1.P.Seq | F           | M00001683B:F11  | CH01COH |
| 223    | 6530    | RTA00002900F.p.04.1.P.Seq | F           | M00005409D:B02  | CH02COH |
| 224    | 1964    | RTA00002903F.c.10.1.P.Seq | F           | M00006885A:F07  | CH02COH |
| 225    | 10547   | RTA00002893F.e.16.1.P.Seq | F           | M00003884A:E12  | CH01COH |
| 226    | 13393   | RTA00002901F.a.06.1.P.Seq | F           | M00005422B:B08  | CH02COH |
| 227    | 14809   | RTA00002915F.e.08.2.P.Seq | F           | M00028773C:C05  | CH08LNL |
| 228    | 46850   | RTA00002907F.e.21.1.P.Seq | F           | M00022158B:B09  | CH03MAH |
| 229    | 5398    | RTA00002911F.i.13.1.P.Seq | F           | M00027004C:C11  | CH04MAL |
| 230    | 27569   | RTA00002910F.l.14.1.P.Seq | F           | M000023007D:D03 | CH03MAH |
| 231    | 26277   | RTA00002898F.g.19.1.P.Seq | F           | M00004347C:A05  | CH01COH |
| 232    | 185914  | RTA00002912F.k.01.1.P.Seq | F           | M00027467A:C07  | CH04MAL |
| 233    | 14274   | RTA00002895F.c.09.1.P.Seq | F           | M00004066D:G10  | CH01COH |
| 234    | 28396   | RTA00002907F.g.02.1.P.Seq | F           | M00022171A:F03  | CH03MAH |
| 235    | 22991   | RTA00002916F.k.07.1.P.Seq | F           | M00032614C:B10  | CH08LNL |
| 236    | 6321    | RTA00002908F.b.17.1.P.Seq | F           | M00022372D:H12  | CH03MAH |
| 237    | 21822   | RTA00002903F.a.17.1.P.Seq | F           | M00006861D:H10  | CH02COH |
| 238    | 8440    | RTA00002923F.k.09.1.P.Seq | F           | M00039302B:E10  | CH09LNL |
| 239    | 14677   | RTA00002905F.c.05.1.P.Seq | F           | M00007975D:F12  | CH03MAH |
| 240    | 135005  | RTA00002902F.d.23.1.P.Seq | F           | M00006592A:A12  | CH02COH |
| 241    | 5091    | RTA00002895F.n.19.1.P.Seq | F           | M00004115A:G12  | CH01COH |
| 242    | 24760   | RTA00002923F.h.18.1.P.Seq | F           | M00039270D:D02  | CH09LNL |
| 243    | 21833   | RTA00002926F.f.22.2.P.Seq | F           | M00040123C:A10  | CH09LNL |
| 244    | 12176   | RTA00002886F.g.18.1.P.Seq | F           | M00001353A:H07  | CH01COH |
| 245    | 14407   | RTA00002901F.c.20.1.P.Seq | F           | M00005452D:E05  | CH02COH |
| 246    | 186319  | RTA00002912F.d.24.1.P.Seq | F           | M00027290C:F06  | CH04MAH |
| 247    | 30135   | RTA00002907F.i.21.2.P.Seq | F           | M00022208C:F08  | CH03MAH |
| 248    | 33142   | RTA00002901F.l.01.1.P.Seq | F           | M00005607B:C04  | CH02COH |
| 249    | 33142   | RTA00002901F.k.24.1.P.Seq | F           | M00005607B:C04  | CH02COH |
| 250    | 16232   | RTA00002900F.n.10.1.P.Seq | F           | M00005387A:B03  | CH02COH |
| 251    | 23954   | RTA00002894F.k.06.1.P.Seq | F           | M00004036B:A11  | CH01COH |
| 252    | 12399   | RTA00002918F.n.10.1.P.Seq | F           | M00032985D:G09  | CH08LNL |
| 253    | 30853   | RTA00002907F.j.16.2.P.Seq | F           | M00022216D:D10  | CH03MAH |
| 254    | 8615    | RTA00002907F.l.17.2.P.Seq | F           | M00022238C:G04  | CH03MAH |
| 255    | 142359  | RTA00002905F.c.10.1.P.Seq | F           | M00007980B:A07  | CH03MAH |
| 256    | 9565    | RTA00002926F.g.08.2.P.Seq | F           | M00040127C:D02  | CH09LNL |
| 257    | 17334   | RTA00002902F.d.03.1.P.Seq | F           | M00006582D:A09  | CH02COH |
| 258    | 12540   | RTA00002886F.f.03.1.P.Seq | F           | M00001349C:B04  | CH01COH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|-------------|-----------------|---------|
| 259    | 17289   | RTA00002926F.f.14.2.P.Seq | F           | M000040118D:C05 | CH09LNL |
| 260    | 46798   | RTA00002907F.i.13.2.P.Seq | F           | M00022202C:C04  | CH03MAH |
| 261    | 7797    | RTA00002905F.b.06.1.P.Seq | F           | M00007953D:F07  | CH03MAH |
| 262    | 1945    | RTA00002896F.p.04.1.P.Seq | F           | M00004201D:C03  | CH01COH |
| 263    | 6084    | RTA00002896F.o.02.1.P.Seq | F           | M00004195A:F07  | CH01COH |
| 264    | 6091    | RTA00002930F.c.03.1.P.Seq | F           | M00042915B:G11  | CH15CON |
| 265    | 186105  | RTA00002930F.c.10.1.P.Seq | F           | M00055430A:A01  | CH15CON |
| 266    | 11341   | RTA00002930F.h.07.1.P.Seq | F           | M00055961C:B10  | CH15CON |
| 267    | 2520    | RTA00002930F.e.10.1.P.Seq | F           | M00055639A:E06  | CH15CON |
| 268    | 136735  | RTA00002903F.k.06.1.P.Seq | F           | M00007006C:C12  | CH02COH |
| 269    | 8336    | RTA00002900F.e.20.1.P.Seq | F           | M00004873B:G04  | CH02COH |
| 270    | 13926   | RTA00002907F.h.19.1.P.Seq | F           | M00022193B:A09  | CH03MAH |
| 271    | 11119   | RTA00002906F.k.01.1.P.Seq | F           | M00022009C:A08  | CH03MAH |
| 272    | 11119   | RTA00002906F.j.24.1.P.Seq | F           | M0002209C:A08   | CH03MAH |
| 273    | 11726   | RTA00002906F.l.07.1.P.Seq | F           | M00022051B:D07  | CH03MAH |
| 274    | 6799    | RTA00002925F.g.21.1.P.Seq | F           | M00039885C:D11  | CH09LNL |
| 275    | 17266   | RTA00002889F.g.09.1.P.Seq | F           | M00001544B:B05  | CH01COH |
| 276    | 9479    | RTA00002924F.g.04.1.P.Seq | F           | M00039560B:G09  | CH09LNL |
| 277    | 185557  | RTA00002912F.j.13.1.P.Seq | F           | M00027457B:E11  | CH04MAL |
| 278    | 27872   | RTA00002906F.e.14.1.P.Seq | F           | M00021923D:H02  | CH03MAH |
| 279    | 15513   | RTA00002924F.g.21.1.P.Seq | F           | M00039617C:A10  | CH09LNL |
| 280    | 4446    | RTA00002891F.m.15.1.P.Seq | F           | M00003773A:F10  | CH01COH |
| 281    | 1681    | RTA00002916F.g.07.1.P.Seq | F           | M00032577D:F01  | CH08LNH |
| 282    | 24243   | RTA00002887F.n.13.1.P.Seq | F           | M00001424D:D02  | CH01COH |
| 283    | 16049   | RTA00002900F.c.11.1.P.Seq | F           | M00004846A:A10  | CH02COH |
| 284    | 186267  | RTA00002910F.h.11.1.P.Seq | F           | M00022924B:A05  | CH03MAH |
| 285    | 4543    | RTA00002925F.h.22.1.P.Seq | F           | M00039895D:C04  | CH09LNL |
| 286    | 6176    | RTA00002914F.d.23.1.P.Seq | F           | M00028188C:H11  | CH08LNH |
| 287    | 29043   | RTA00002906F.h.17.1.P.Seq | F           | M00021974D:F01  | CH03MAH |
| 288    | 696     | RTA00002922F.o.15.1.P.Seq | F           | M00039143A:F04  | CH09LNL |
| 289    | 7225    | RTA00002891F.l.22.1.P.Seq | F           | M00003770C:A10  | CH01COH |
| 290    | 25609   | RTA00002899F.h.15.1.P.Seq | F           | M00004608A:H04  | CH01COH |
| 291    | 6295    | RTA00002922F.o.24.1.P.Seq | F           | M00039146B:G04  | CH09LNL |
| 292    | 186319  | RTA00002912F.e.01.1.P.Seq | F           | M00027290C:F06  | CH04MAL |
| 293    | 4539    | RTA00002889F.d.04.1.P.Seq | F           | M00001534C:E07  | CH01COH |
| 294    | 17841   | RTA00002891F.m.06.1.P.Seq | F           | M00003771D:A03  | CH01COH |
| 295    | 13720   | RTA00002924F.c.05.1.P.Seq | F           | M000039430A:E04 | CH09LNL |
| 296    | 7300    | RTA00002925F.a.14.1.P.Seq | F           | M00039806B:D05  | CH09LNL |
| 297    | 186280  | RTA00002912F.f.13.1.P.Seq | F           | M00027316C:C03  | CH04MAL |
| 298    | 185585  | RTA00002912F.n.04.1.P.Seq | F           | M00027569A:E05  | CH04MAL |
| 299    | 3447    | RTA00002900F.l.11.1.P.Seq | F           | M00005364B:E10  | CH02COH |
| 300    | 14487   | RTA00002889F.f.19.1.P.Seq | F           | M00001542B:F09  | CH01COH |
| 301    | 5338    | RTA00002901F.g.16.1.P.Seq | F           | M00005505A:F01  | CH02COH |
| 302    | 7766    | RTA00002917F.e.07.1.P.Seq | F           | M00032700A:E09  | CH08LNH |
| 303    | 7450    | RTA00002899F.g.09.1.P.Seq | F           | M00004509D:C06  | CH01COH |
| 304    | 15369   | RTA00002908F.m.18.1.P.Seq | F           | M00022494D:A05  | CH03MAH |
| 305    | 4954    | RTA00002919F.i.17.1.P.Seq | F           | M00033150C:A11  | CH08LNH |
| 306    | 17189   | RTA00002900F.j.11.1.P.Seq | F           | M00005333D:D08  | CH02COH |
| 307    | 186561  | RTA00002912F.m.23.1.P.Seq | F           | M00027549C:G03  | CH04MAL |
| 308    | 44645   | RTA00002896F.h.22.1.P.Seq | F           | M00004165C:A11  | CH01COH |
| 309    | 11404   | RTA00002924F.a.24.1.P.Seq | F           | M00039413C:E06  | CH09LNL |
| 310    | 38212   | RTA00002893F.m.22.1.P.Seq | F           | M00003942A:D01  | CH01COH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|-------------|-----------------|---------|
| 311    | 27099   | RTA00002890F.m.09.1.P.Seq | F           | M00001648A:D10  | CH01COH |
| 312    | 25041   | RTA00002890F.p.12.1.P.Seq | F           | M00001661D:F06  | CH01COH |
| 313    | 185938  | RTA00002911F.p.01.1.P.Seq | F           | M00027173C:E11  | CH04MAL |
| 314    | 9414    | RTA00002908F.o.06.1.P.Seq | F           | M00022509B:D11  | CH03MAH |
| 315    | 185707  | RTA00002911F.o.19.1.P.Seq | F           | M00027167C:B10  | CH04MAL |
| 316    | 185499  | RTA00002912F.n.19.1.P.Seq | F           | M00027586A:C09  | CH04MAL |
| 317    | 25704   | RTA00002912F.n.22.1.P.Seq | F           | M00027589B:G07  | CH04MAL |
| 318    | 21068   | RTA00002896F.h.18.1.P.Seq | F           | M00004164B:E12  | CH01COH |
| 319    | 13440   | RTA00002917F.e.18.1.P.Seq | F           | M00032711B:F01  | CH08LNH |
| 320    | 3907    | RTA00002923F.i.18.1.P.Seq | F           | M00039285B:G04  | CH09LNL |
| 321    | 21391   | RTA00002896F.g.03.1.P.Seq | F           | M00004158D:E08  | CH01COH |
| 322    | 6755    | RTA00002918F.1.01.1.P.Seq | F           | M00032944A:B07  | CH08LNH |
| 323    | 155939  | RTA00002907F.i.23.1.P.Seq | F           | M00022218B:B12  | CH03MAH |
| 324    | 8100    | RTA00002896F.g.21.1.P.Seq | F           | M00004160D:F06  | CH01COH |
| 325    | 4785    | RTA00002919F.i.18.1.P.Seq | F           | M00033183B:F10  | CH08LNH |
| 326    | 14947   | RTA00002902F.k.23.1.P.Seq | F           | M00006743A:D04  | CH02COH |
| 327    | 8295    | RTA00002903F.k.23.1.P.Seq | F           | M00007031A:E02  | CH02COH |
| 328    | 156277  | RTA00002907F.1.13.1.P.Seq | F           | M00022237D:D06  | CH03MAH |
| 329    | 22751   | RTA00002897F.1.15.1.P.Seq | F           | M00004282C:A12  | CH01COH |
| 330    | 7869    | RTA00002917F.i.15.1.P.Seq | F           | M00032749D:G03  | CH08LNH |
| 331    | 156009  | RTA00002907F.k.05.1.P.Seq | F           | M0002222A:A07   | CH03MAH |
| 332    | 9453    | RTA00002907F.k.21.1.P.Seq | F           | M00022228B:B11  | CH03MAH |
| 333    | 186052  | RTA00002912F.h.08.1.P.Seq | F           | M00027364B:E12  | CH04MAL |
| 334    | 669     | RTA00002917F.f.22.1.P.Seq | F           | M00032723D:H02  | CH08LNH |
| 335    | 11609   | RTA00002899F.f.23.1.P.Seq | F           | M00004507D:E03  | CH01COH |
| 336    | 186075  | RTA00002911F.k.19.1.P.Seq | F           | M00027057C:D10  | CH04MAL |
| 337    | 935     | RTA00002911F.1.20.1.P.Seq | F           | M00027081A:A08  | CH04MAL |
| 338    | 11430   | RTA00002892F.e.07.2.P.Seq | F           | M00003808B:E07  | CH01COH |
| 339    | 185938  | RTA00002911F.o.24.1.P.Seq | F           | M00027173C:E11  | CH04MAL |
| 340    | 12394   | RTA00002915F.m.15.2.P.Seq | F           | M00032497D:B10  | CH08LNH |
| 341    | 186588  | RTA00002911F.1.03.1.P.Seq | F           | M00027064B:D06  | CH04MAL |
| 342    | 23174   | RTA00002909F.e.17.1.P.Seq | F           | M00022600D:B05  | CH03MAH |
| 343    | 4727    | RTA00002905F.g.19.1.P.Seq | F           | M00008059D:B08  | CH03MAH |
| 344    | 17048   | RTA00002887F.1.10.1.P.Seq | F           | M00001416B:A05  | CH01COH |
| 345    | 2354    | RTA00002916F.o.03.1.P.Seq | F           | M00032645D:C01  | CH08LNH |
| 346    | 1987    | RTA00002894F.a.13.1.P.Seq | F           | M00003974D:E02  | CH01COH |
| 347    | 24483   | RTA00002897F.i.21.1.P.Seq | F           | M00004269A:G11  | CH01COH |
| 348    | 33337   | RTA00002896F.f.08.1.P.Seq | F           | M00004155A:H03  | CH01COH |
| 349    | 11641   | RTA00002916F.m.19.1.P.Seq | F           | M00032637A:F09  | CH08LNH |
| 350    | 10307   | RTA00002910F.1.01.1.P.Seq | F           | M00022995C:G07  | CH03MAH |
| 351    | 20388   | RTA00002906F.a.04.1.P.Seq | F           | M00021700D:H03  | CH03MAH |
| 352    | 24687   | RTA00002903F.m.02.1.P.Seq | F           | M00007048B:E11  | CH02COH |
| 353    | 10414   | RTA00002919F.n.19.1.P.Seq | F           | M00033232B:C08  | CH08LNH |
| 354    | 11058   | RTA00002892F.h.16.2.P.Seq | F           | M00003820B:F11  | CH01COH |
| 355    | 6574    | RTA00002917F.o.17.1.P.Seq | F           | M00032797D:D08  | CH08LNH |
| 356    | 18782   | RTA00002905F.f.07.1.P.Seq | F           | M00008021C:G12  | CH03MAH |
| 357    | 35896   | RTA00002896F.d.04.1.P.Seq | F           | M00004146C:B04  | CH01COH |
| 358    | 3518    | RTA00002930F.i.10.1.P.Seq | F           | M00056217D:E10  | CH15CON |
| 359    | 8820    | RTA00002915F.f.17.2.P.Seq | F           | M000028782A:F01 | CH08LNH |
| 360    | 10208   | RTA00002897F.h.08.1.P.Seq | F           | M0004251D:D03   | CH01COH |
| 361    | 2089    | RTA00002896F.g.14.1.P.Seq | F           | M00004159D:F12  | CH01COH |
| 362    | 170919  | RTA00002909F.p.03.1.P.Seq | F           | M00022727A:G01  | CH03MAH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|-------------|-----------------|---------|
| 363    | 8727    | RTA00002917F.o.02.1.P.Seq | F           | M000032791B:H11 | CH08LNH |
| 364    | 33184   | RTA00002898F.d.08.1.P.Seq | F           | M00004324A:D10  | CH01COH |
| 365    | 27973   | RTA00002905F.g.13.1.P.Seq | F           | M00008055D:G03  | CH03MAH |
| 366    | 15835   | RTA00002897F.k.13.1.P.Seq | F           | M00004278C:B10  | CH01COH |
| 367    | 10273   | RTA00002903F.n.03.1.P.Seq | F           | M00007081B:E09  | CH02COH |
| 368    | 2832    | RTA00002899F.f.03.1.P.Seq | F           | M00004502A:D12  | CH01COH |
| 369    | 32022   | RTA00002903F.m.12.1.P.Seq | F           | M00007060D:G07  | CH02COH |
| 370    | 68176   | RTA00002893F.g.11.1.P.Seq | F           | M00003898C:A01  | CH01COH |
| 371    | 29378   | RTA00002915F.n.14.1.P.Seq | F           | M00032508A:E03  | CH08LNH |
| 372    | 23235   | RTA00002925F.k.02.1.P.Seq | F           | M00039929B:E06  | CH09LNL |
| 373    | 12111   | RTA00002895F.o.17.1.P.Seq | F           | M00004122C:D01  | CH01COH |
| 374    | 5737    | RTA00002924F.k.02.1.P.Seq | F           | M00039672C:D05  | CH09LNL |
| 375    | 72475   | RTA00002915F.l.15.1.P.Seq | F           | M00032490D:E08  | CH08LNH |
| 376    | 7027    | RTA00002907F.o.01.1.P.Seq | F           | M00022264A:B02  | CH03MAH |
| 377    | 17165   | RTA00002903F.d.19.1.P.Seq | F           | M00006907A:C09  | CH02COH |
| 378    | 26446   | RTA00002894F.m.17.1.P.Seq | F           | M00004047C:B09  | CH01COH |
| 379    | 6755    | RTA00002918F.k.24.1.P.Seq | F           | M00032944A:B07  | CH08LNH |
| 380    | 9336    | RTA00002909F.n.02.1.P.Seq | F           | M00022703D:B11  | CH03MAH |
| 381    | 6960    | RTA00002916F.o.08.1.P.Seq | F           | M00032647B:F06  | CH08LNH |
| 382    | 472     | RTA00002911F.g.01.1.P.Seq | F           | M00026936D:C07  | CH04MAL |
| 383    | 9460    | RTA00002908F.c.03.1.P.Seq | F           | M00022376D:D05  | CH03MAH |
| 384    | 10307   | RTA00002910F.k.24.1.P.Seq | F           | M00022995C:G07  | CH03MAH |
| 385    | 4623    | RTA00002923F.d.22.1.P.Seq | F           | M00039222B:A04  | CH09LNL |
| 386    | 141167  | RTA00002905F.c.09.1.P.Seq | F           | M00007980A:B01  | CH03MAH |
| 387    | 34011   | RTA00002898F.m.10.1.P.Seq | F           | M00004385C:H12  | CH01COH |
| 388    | 5965    | RTA00002915F.a.07.1.P.Seq | F           | M00028620C:C07  | CH08LNH |
| 389    | 12336   | RTA00002915F.g.04.1.P.Seq | F           | M00028784A:D12  | CH08LNH |
| 390    | 36492   | RTA00002893F.f.18.1.P.Seq | F           | M00003891B:H02  | CH01COH |
| 391    | 29803   | RTA00002908F.k.06.1.P.Seq | F           | M00022467D:B03  | CH03MAH |
| 392    | 4420    | RTA00002920F.a.15.1.P.Seq | F           | M00033326B:B05  | CH08LNH |
| 393    | 15097   | RTA00002923F.b.06.1.P.Seq | F           | M00039175A:F01  | CH09LNL |
| 394    | 19133   | RTA00002894F.g.03.1.P.Seq | F           | M00003993C:D07  | CH01COH |
| 395    | 9810    | RTA00002905F.c.03.1.P.Seq | F           | M00007975C:A10  | CH03MAH |
| 396    | 31562   | RTA00002897F.a.09.1.P.Seq | F           | M00004210A:A03  | CH01COH |
| 397    | 1499    | RTA00002912F.k.12.1.P.Seq | F           | M00027475D:A01  | CH04MAL |
| 398    | 29531   | RTA00002907F.o.15.1.P.Seq | F           | M00022265A:F11  | CH03MAH |
| 399    | 4287    | RTA00002918F.i.20.1.P.Seq | F           | M00032928C:D02  | CH08LNH |
| 400    | 28660   | RTA00002905F.p.11.1.P.Seq | F           | M00021690A:C03  | CH03MAH |
| 401    | 4596    | RTA00002898F.f.21.1.P.Seq | F           | M00004341C:E05  | CH01COH |
| 402    | 21774   | RTA00002898F.c.20.1.P.Seq | F           | M00004322B:D03  | CH01COH |
| 403    | 5611    | RTA00002915F.e.12.1.P.Seq | F           | M000028774D:E10 | CH08LNH |
| 404    | 7030    | RTA00002894F.l.13.1.P.Seq | F           | M00004042B:A11  | CH01COH |
| 405    | 11736   | RTA00002898F.e.09.1.P.Seq | F           | M00004330A:A01  | CH01COH |
| 406    | 94732   | RTA00002910F.e.17.1.P.Seq | F           | M00022856D:A07  | CH03MAH |
| 407    | 30283   | RTA00002923F.g.19.1.P.Seq | F           | M00039255D:B01  | CH09LNL |
| 408    | 129779  | RTA00002904F.a.18.1.P.Seq | F           | M00007155C:D07  | CH02COH |
| 409    | 4635    | RTA00002900F.i.21.1.P.Seq | F           | M00005349C:C02  | CH02COH |
| 410    | 5879    | RTA00002893F.f.08.1.P.Seq | F           | M00003888B:F09  | CH01COH |
| 411    | 119206  | RTA00002905F.m.16.1.P.Seq | F           | M00021650D:A11  | CH03MAH |
| 412    | 6946    | RTA00002930F.g.19.2.P.Seq | F           | M00055880B:H10  | CH15CON |
| 413    | 42462   | RTA00002902F.f.12.1.P.Seq | F           | M00006631D:D02  | CH02COH |
| 414    | 24285   | RTA00002896F.m.17.1.P.Seq | F           | M00004189A:C12  | CH01COH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|-------------|----------------|---------|
| 415    | 13769   | RTA00002901F.a.17.1.P.Seq | F           | M00005423C:D07 | CH02COH |
| 416    | 17039   | RTA00002896F.i.14.1.P.Seq | F           | M00004169A:E04 | CH01COH |
| 417    | 14397   | RTA00002896F.i.11.1.P.Seq | F           | M00004172D:B12 | CH01COH |
| 418    | 14351   | RTA00002888F.c.21.1.P.Seq | F           | M00001444C:D11 | CH01COH |
| 419    | 5579    | RTA00002893F.j.11.1.P.Seq | F           | M00003914A:A08 | CH01COH |
| 420    | 24186   | RTA00002914F.n.02.1.P.Seq | F           | M00028366B:B08 | CH08LNH |
| 421    | 11433   | RTA00002921F.c.06.1.P.Seq | F           | M00033342B:F03 | CH09LNL |
| 422    | 186635  | RTA00002911F.f.06.1.P.Seq | F           | M00026907D:E07 | CH04MAL |
| 423    | 5955    | RTA00002915F.d.18.1.P.Seq | F           | M00028771A:E02 | CH08LNH |
| 424    | 22053   | RTA00002894F.k.09.1.P.Seq | F           | M00004036D:C12 | CH01COH |
| 425    | 9259    | RTA00002918F.b.09.1.P.Seq | F           | M00032830D:D02 | CH08LNH |
| 426    | 25437   | RTA00002905F.o.23.1.P.Seq | F           | M00021681C:C09 | CH03MAH |
| 427    | 8488    | RTA00002916F.i.02.1.P.Seq | F           | M00032590B:H01 | CH08LNH |
| 428    | 4884    | RTA00002919F.o.12.1.P.Seq | F           | M00033248D:H11 | CH08LNH |
| 429    | 9804    | RTA00002915F.c.19.1.P.Seq | F           | M00028764B:D03 | CH08LNH |
| 430    | 179954  | RTA00002910F.i.04.1.P.Seq | F           | M00022964A:B03 | CH03MAH |
| 431    | 186532  | RTA00002912F.a.01.1.P.Seq | F           | M00027189C:B10 | CH04MAL |
| 432    | 11015   | RTA00002894F.i.15.1.P.Seq | F           | M00004029D:A01 | CH01COH |
| 433    | 8824    | RTA00002903F.b.17.1.P.Seq | F           | M00006877B:C09 | CH02COH |
| 434    | 4063    | RTA00002916F.k.01.1.P.Seq | F           | M00032613A:E11 | CH08LNH |
| 435    | 7964    | RTA00002896F.i.18.1.P.Seq | F           | M00004170A:F03 | CH01COH |
| 436    | 9238    | RTA00002915F.i.20.1.P.Seq | F           | M00032473B:A03 | CH08LNH |
| 437    | 2841    | RTA00002914F.f.15.1.P.Seq | F           | M00028196A:G03 | CH08LNH |
| 438    | 11203   | RTA00002886F.p.16.1.P.Seq | F           | M00001382D:H08 | CH01COH |
| 439    | 8800    | RTA00002888F.c.20.1.P.Seq | F           | M00001444B:E04 | CH01COH |
| 440    | 3224    | RTA00002916F.d.23.1.P.Seq | F           | M00032556D:A03 | CH08LNH |
| 441    | 95423   | RTA00002909F.k.24.1.P.Seq | F           | M00022674C:H08 | CH03MAH |
| 442    | 7911    | RTA00002926F.c.11.2.P.Seq | F           | M00040079D:D09 | CH09LNL |
| 443    | 88052   | RTA00002925F.p.11.1.P.Seq | F           | M00040041D:F01 | CH09LNL |
| 444    | 32736   | RTA00002900F.l.20.1.P.Seq | F           | M00005367D:A11 | CH02COH |
| 445    | 20811   | RTA00002896F.n.14.1.P.Seq | F           | M00004192C:B06 | CH01COH |
| 446    | 12856   | RTA00002908F.b.07.1.P.Seq | F           | M00022368A:B11 | CH03MAH |
| 447    | 12190   | RTA00002899F.b.10.1.P.Seq | F           | M0004430B:B10  | CH01COH |
| 448    | 10546   | RTA00002901F.o.08.1.P.Seq | F           | M0005689C:B02  | CH02COH |
| 449    | 21041   | RTA00002898F.k.08.1.P.Seq | F           | M00004372A:E12 | CH01COH |
| 450    | 16484   | RTA00002894F.c.04.1.P.Seq | F           | M00003979B:A04 | CH01COH |
| 451    | 7741    | RTA00002894F.i.08.1.P.Seq | F           | M00004028B:F10 | CH01COH |
| 452    | 14921   | RTA00002926F.c.15.2.P.Seq | F           | M00040081C:E02 | CH09LNL |
| 453    | 17571   | RTA00002900F.m.16.1.P.Seq | F           | M00005375D:A10 | CH02COH |
| 454    | 46881   | RTA00002901F.l.20.1.P.Seq | F           | M00005622A:H02 | CH02COH |
| 455    | 21533   | RTA00002898F.j.10.1.P.Seq | F           | M00004365C:C09 | CH01COH |
| 456    | 19010   | RTA00002916F.k.08.1.P.Seq | F           | M00032614D:D08 | CH08LNH |
| 457    | 48768   | RTA00002886F.n.01.1.P.Seq | F           | M00001374C:B10 | CH01COH |
| 458    | 7515    | RTA00002892F.p.22.2.P.Seq | F           | M00003855C:F02 | CH01COH |
| 459    | 17326   | RTA00002898F.h.02.1.P.Seq | F           | M00004350A:A04 | CH01COH |
| 460    | 3902    | RTA00002901F.d.17.1.P.Seq | F           | M00005460D:C11 | CH02COH |
| 461    | 12400   | RTA00002901F.d.18.1.P.Seq | F           | M00005461A:D12 | CH02COH |
| 462    | 186543  | RTA00002912F.a.06.1.P.Seq | F           | M00027193C:A07 | CH04MAL |
| 463    | 4063    | RTA00002916F.j.24.1.P.Seq | F           | M00032613A:E11 | CH08LNH |
| 464    | 6267    | RTA00002910F.d.20.1.P.Seq | F           | M00022835C:A09 | CH03MAH |
| 465    | 21349   | RTA00002901F.c.04.1.P.Seq | F           | M00005445D:F11 | CH02COH |
| 466    | 1123    | RTA00002894F.i.24.1.P.Seq | F           | M00004031C:G06 | CH01COH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID        | LIBRARY   |
|--------|---------|---------------------------|--------------|-----------------|-----------|
| 467    | 4401    | RTA00002918F.m.18.1.P.Seq | F            | M000032979D:C11 | CH08LNH   |
| 468    | 15255   | RTA00002925F.p.10.1.P.Seq | F            | M00040041A:G08  | CH09LNL   |
| 469    | 10991   | RTA00002933F.a.15.1.P.Seq | F            | M00043077C:G10  | CH19COP   |
| 470    | 48768   | RTA00002886F.m.24.1.P.Seq | F            | M00001374C:B10  | CH01COH   |
| 471    | 20406   | RTA00002900F.c.20.1.P.Seq | F            | M00004852D:C06  | CH02COH   |
| 472    | 39784   | RTA00002886F.g.05.1.P.Seq | F            | M00001352B:B02  | CH01COH   |
| 473    | 36567   | RTA00002886F.n.06.1.P.Seq | F            | M00001375B:D04  | CH01COH   |
| 474    | 14817   | RTA00002902F.a.18.1.P.Seq | F            | M00005771D:C02  | CH02COH   |
| 475    | 156277  | RTA00002907F.1.13.2.P.Seq | F            | M00022237D:D06  | CH03MAH   |
| 476    | 6898    | RTA00002907F.a.22.1.P.Seq | F            | M00022104A:G08  | CH03MAH   |
| 477    | 17376   | RTA00002902F.c.03.1.P.Seq | F            | M00005819D:F09  | CH02COH   |
| 478    | 186535  | RTA00002912F.d.12.1.P.Seq | F            | M00027270A:D04  | CH04MAL   |
| 479    | 91616   | RTA00002910F.b.24.1.P.Seq | F            | M00022812A:G01  | CH03MAH   |
| 480    | 91616   | RTA00002910F.c.01.1.P.Seq | F            | M00022812A:G01  | CH03MAH   |
| 481    | 6993    | RTA00002896F.i.12.1.P.Seq | F            | M00004172D:F04  | CH01COH   |
| 482    | 12443   | RTA00002916F.a.20.1.P.Seq | F            | M00032534B:E12  | CH08LNH   |
| 483    | 28585   | RTA00002901F.i.16.1.P.Seq | F            | M00005570A:D05  | CH02COH   |
| 484    | 9453    | RTA00002907F.k.21.2.P.Seq | F            | M00022228B:B11  | CH03MAH   |
| 485    | 156009  | RTA00002907F.k.05.2.P.Seq | F            | M00022220A:A07  | CH03MAH   |
| 486    | 5958    | RTA00002908F.n.22.2.P.Seq | F            | M00022507C:C08  | CH03MAH   |
| 487    | 155939  | RTA00002907F.i.23.2.P.Seq | F            | M00022218B:B12  | CH03MAH   |
| 488    | 16695   | RTA00002886F.g.22.1.P.Seq | F            | M00001353D:E05  | CH01COH   |
| 489    | 10118   | RTA00002886F.h.18.1.P.Seq | F            | M00001356D:E06  | CH01COH   |
| 490    | 13288   | RTA00002930F.b.21.1.P.Seq | F            | M00042891C:G08  | CH15CON   |
| 491    | 3210    | RTA00002910F.h.22.1.P.Seq | F            | M00022945A:H09  | CH03MAH   |
| 492    | 15014   | RTA00002934F.a.18.1.P.Seq | F            | M00043528A:E11  | CH20COHLV |
| 493    | 22087   | RTA00002890F.i.19.1.P.Seq | F            | M00001624A:C01  | CH01COH   |
| 494    | 31948   | RTA00002908F.i.12.1.P.Seq | F            | M00022454C:B08  | CH03MAH   |
| 495    | 11593   | RTA00002906F.p.21.1.P.Seq | F            | M00022094B:G02  | CH03MAH   |
| 496    | 3131    | RTA00002908F.m.17.1.P.Seq | F            | M00022494B:D06  | CH03MAH   |
| 497    | 151263  | RTA00002906F.i.21.1.P.Seq | F            | M00021991D:F09  | CH03MAH   |
| 498    | 177542  | RTA00002910F.h.23.1.P.Seq | F            | M00022945B:F11  | CH03MAH   |
| 499    | 9738    | RTA00002924F.f.23.1.P.Seq | F            | M00039559B:C07  | CH09LNL   |
| 500    | 15313   | RTA00002925F.f.05.1.P.Seq | F            | M00039865A:C09  | CH09LNL   |
| 501    | 19724   | RTA00002906F.p.05.1.P.Seq | F            | M00022088D:E10  | CH03MAH   |
| 502    | 10731   | RTA00002893F.m.11.1.P.Seq | F            | M00003938C:A05  | CH01COH   |
| 503    | 10257   | RTA00002901F.h.09.1.P.Seq | F            | M00005512B:H01  | CH02COH   |
| 504    | 186468  | RTA00002913F.b.18.1.P.Seq | F            | M000027746A:D06 | CH04MAL   |
| 505    | 14736   | RTA00002908F.g.22.1.P.Seq | F            | M000022436C:F11 | CH03MAH   |
| 506    | 33267   | RTA00002889F.h.14.1.P.Seq | F            | M00001548B:D06  | CH01COH   |
| 507    | 7719    | RTA00002908F.e.11.1.P.Seq | F            | M00022403C:E12  | CH03MAH   |
| 508    | 185539  | RTA00002913F.b.03.1.P.Seq | F            | M00027717C:C06  | CH04MAL   |
| 509    | 14825   | RTA00002924F.f.19.1.P.Seq | F            | M00039556C:G05  | CH09LNL   |
| 510    | 3917    | RTA00002906F.p.15.1.P.Seq | F            | M00022092D:A11  | CH03MAH   |
| 511    | 18718   | RTA00002895F.h.05.1.P.Seq | F            | M00004085B:H02  | CH01COH   |
| 512    | 186762  | RTA00002918F.b.11.1.P.Seq | F            | M00032831A:C07  | CH08LNH   |
| 513    | 2732    | RTA00002925F.i.07.1.P.Seq | F            | M00039900B:G04  | CH09LNL   |
| 514    | 7684    | RTA00002924F.i.17.1.P.Seq | F            | M00039668C:F01  | CH09LNL   |
| 515    | 6852    | RTA00002922F.c.22.1.P.Seq | F            | M00039001A:B10  | CH09LNL   |
| 516    | 1422    | RTA00002924F.c.09.1.P.Seq | F            | M00039433C:E03  | CH09LNL   |
| 517    | 5560    | RTA00002917F.i.01.1.P.Seq | F            | M00032734C:C05  | CH08LNH   |
| 518    | 48734   | RTA00002908F.i.23.1.P.Seq | F            | M00022487C:C02  | CH03MAH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|--------------|-----------------|---------|
| 519    | 10486   | RTA00002899F.a.07.1.P.Seq | F            | M00004509B:R10  | CH01COH |
| 520    | 33514   | RTA00002890F.i.03.1.P.Seq | F            | M00001626A:D07  | CH01COH |
| 521    | 5821    | RTA00002917F.m.01.1.P.Seq | F            | M00032772D:D03  | CH08LNL |
| 522    | 5821    | RTA00002917F.l.24.1.P.Seq | F            | M00032772D:D03  | CH08LNL |
| 523    | 21940   | RTA00002896F.a.03.1.P.Seq | F            | M00004134A:A8   | CH01COH |
| 524    | 185724  | RTA00002912F.m.08.1.P.Seq | F            | M00027523A:H05  | CH04MAL |
| 525    | 182887  | RTA00002910F.k.21.1.P.Seq | F            | M00022992A:H06  | CH03MAH |
| 526    | 21346   | RTA00002901F.g.24.1.P.Seq | F            | M00005507B:A03  | CH02COH |
| 527    | 5501    | RTA00002887F.n.12.1.P.Seq | F            | M00001424B:H06  | CH01COH |
| 528    | 13961   | RTA00002892F.i.14.1.P.Seq | F            | M00003828A:D11  | CH01COH |
| 529    | 16784   | RTA00002886F.a.09.1.P.Seq | F            | M00001338D:D01  | CH01COH |
| 530    | 17628   | RTA00002916F.f.10.1.P.Seq | F            | M00032568B:F08  | CH08LNL |
| 531    | 3304    | RTA00002898F.d.05.1.P.Seq | F            | M00004324A:B03  | CH01COH |
| 532    | 14895   | RTA00002901F.g.14.1.P.Seq | F            | M00005504C:F12  | CH02COH |
| 533    | 16036   | RTA00002891F.k.09.1.P.Seq | F            | M00003763B:B10  | CH01COH |
| 534    | 23877   | RTA00002891F.k.15.1.P.Seq | F            | M00003764B:F11  | CH01COH |
| 535    | 186784  | RTA00002930F.i.17.1.P.Seq | F            | M00056105A:D06  | CH15CON |
| 536    | 13591   | RTA00002901F.f.15.1.P.Seq | F            | M00005485C:H04  | CH02COH |
| 537    | 17916   | RTA00002906F.p.08.1.P.Seq | F            | M00022090B:A10  | CH03MAH |
| 538    | 40594   | RTA00002897F.i.15.1.P.Seq | F            | M00004266B:F07  | CH01COH |
| 539    | 9677    | RTA00002925F.i.21.1.P.Seq | F            | M00039915B:E08  | CH09LNL |
| 540    | 7736    | RTA00002887F.e.03.1.P.Seq | F            | M00001393B:C03  | CH01COH |
| 541    | 2474    | RTA00002917F.e.15.1.P.Seq | F            | M00032707D:F08  | CH08LNL |
| 542    | 23810   | RTA00002892F.i.06.1.P.Seq | F            | M00003822D:A02  | CH01COH |
| 543    | 24633   | RTA00002907F.i.19.1.P.Seq | F            | M00022208B:D03  | CH03MAH |
| 544    | 72081   | RTA00002925F.k.03.1.P.Seq | F            | M00039929D:H10  | CH09LNL |
| 545    | 5991    | RTA00002916F.i.17.1.P.Seq | F            | M00032597A:H02  | CH08LNL |
| 546    | 14596   | RTA00002911F.n.15.1.P.Seq | F            | M00027131A:B03  | CH04MAL |
| 547    | 6923    | RTA00002896F.d.01.1.P.Seq | F            | M00004146B:E08  | CH01COH |
| 548    | 6923    | RTA00002896F.c.24.1.P.Seq | F            | M00004146B:E08  | CH01COH |
| 549    | 21851   | RTA00002887F.d.09.1.P.Seq | F            | M00001391D:D03  | CH01COH |
| 550    | 3935    | RTA00002925F.i.08.1.P.Seq | F            | M00039921C:H11  | CH09LNL |
| 551    | 13328   | RTA00002909F.h.08.1.P.Seq | F            | M000022634B:H09 | CH03MAH |
| 552    | 2492    | RTA00002887F.e.11.1.P.Seq | F            | M00001393D:E02  | CH01COH |
| 553    | 11960   | RTA00002917F.b.03.1.P.Seq | F            | M00032671B:D08  | CH08LNL |
| 554    | 186084  | RTA00002912F.f.18.1.P.Seq | F            | M00027319D:F07  | CH04MAL |
| 555    | 13644   | RTA00002925Fa.09.1.P.Seq  | F            | M00039805B:B06  | CH09LNL |
| 556    | 5707    | RTA00002909F.k.13.1.P.Seq | F            | M00022672C:H04  | CH03MAH |
| 557    | 95700   | RTA00002911F.p.14.1.P.Seq | F            | M00027182B:G06  | CH04MAL |
| 558    | 342     | RTA00002922F.i.23.1.P.Seq | F            | M00039076D:G04  | CH09LNL |
| 559    | 8481    | RTA00002887F.c.12.1.P.Seq | F            | M00001389D:D06  | CH01COH |
| 560    | 12575   | RTA00002916F.i.12.1.P.Seq | F            | M000032594C:F05 | CH08LNL |
| 561    | 40712   | RTA00002921F.d.08.1.P.Seq | F            | M00033359C:H05  | CH09LNL |
| 562    | 10768   | RTA00002886F.d.24.1.P.Seq | F            | M00001346B:G11  | CH01COH |
| 563    | 38781   | RTA00002889F.k.23.1.P.Seq | F            | M00001559A:H09  | CH01COH |
| 564    | 8790    | RTA00002888F.g.08.1.P.Seq | F            | M00001461D:C10  | CH01COH |
| 565    | 10167   | RTA00002916F.k.22.1.P.Seq | F            | M00032621A:F11  | CH08LNL |
| 566    | 13706   | RTA00002905F.e.21.1.P.Seq | F            | M00008019B:A01  | CH03MAH |
| 567    | 124172  | RTA00002900F.a.09.1.P.Seq | F            | M00004824A:D12  | CH02COH |
| 568    | 92126   | RTA00002910F.g.12.1.P.Seq | F            | M00022904C:D04  | CH03MAH |
| 569    | 5830    | RTA00002916F.i.09.1.P.Seq | F            | M00032605B:D09  | CH08LNL |
| 570    | 15154   | RTA00002886F.p.13.1.P.Seq | F            | M00001382D:A07  | CH01COH |



| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID        | LIBRARY |
|--------|---------|---------------------------|--------------|-----------------|---------|
| 571    | 75813   | RTA00002910F.i.12.1.P.Seq | F            | M000022957A:B07 | CH03MAH |
| 572    | 17268   | RTA00002886F.d.07.1.P.Seq | F            | M00001344D:E08  | CH01COH |
| 573    | 13684   | RTA00002915F.i.09.1.P.Seq | F            | M00031485B:G05  | CH08LNH |
| 574    | 13460   | RTA00002898F.f.19.1.P.Seq | F            | M00004341C:A09  | CH01COH |
| 575    | 25115   | RTA00002919F.p.18.1.P.Seq | F            | M00033311B:G10  | CH08LNH |
| 576    | 19949   | RTA00002905F.e.17.1.P.Seq | F            | M00008016B:E09  | CH03MAH |
| 577    | 24266   | RTA00002917F.k.06.1.P.Seq | F            | M00032759A:A03  | CH08LNH |
| 578    | 8243    | RTA00002901F.o.17.1.P.Seq | F            | M00005703B:E03  | CH02COH |
| 579    | 12576   | RTA00002900F.k.23.1.P.Seq | F            | M00005359B:B08  | CH02COH |
| 580    | 28531   | RTA00002909F.c.04.1.P.Seq | F            | M00022559D:G10  | CH03MAH |
| 581    | 15153   | RTA00002894F.o.21.1.P.Seq | F            | M00004054A:D03  | CH01COH |
| 582    | 9498    | RTA00002894F.e.04.1.P.Seq | F            | M00003985D:B02  | CH01COH |
| 583    | 48140   | RTA00002914F.h.13.1.P.Seq | F            | M00028211A:F10  | CH08LNH |
| 584    | 7626    | RTA00002895F.b.04.1.P.Seq | F            | M00004061B:E05  | CH01COH |
| 585    | 22668   | RTA00002896F.p.17.1.P.Seq | F            | M00004204C:H08  | CH01COH |
| 586    | 45691   | RTA00002908F.a.11.1.P.Seq | F            | M00022305A:B04  | CH03MAH |
| 587    | 30429   | RTA00002904F.a.19.1.P.Seq | F            | M00007155D:C09  | CH02COH |
| 588    | 46969   | RTA00002909F.g.02.1.P.Seq | F            | M00022618C:E04  | CH03MAH |
| 589    | 44030   | RTA00002900F.o.23.1.P.Seq | F            | M00005405C:D01  | CH02COH |
| 590    | 142548  | RTA00002905F.h.10.1.P.Seq | F            | M00008073A:D01  | CH03MAH |
| 591    | 18455   | RTA00002905F.g.18.1.P.Seq | F            | M00008059B:F08  | CH03MAH |
| 592    | 7501    | RTA00002894F.g.05.1.P.Seq | F            | M00003993D:B03  | CH01COH |
| 593    | 7280    | RTA00002893F.n.22.1.P.Seq | F            | M00003959D:A05  | CH01COH |
| 594    | 19339   | RTA00002898F.l.12.1.P.Seq | F            | M00004376D:A12  | CH01COH |
| 595    | 30194   | RTA00002922F.k.05.1.P.Seq | F            | M00039100A:G04  | CH09LNL |
| 596    | 32650   | RTA00002911F.i.05.1.P.Seq | F            | M00026994D:D07  | CH04MAL |
| 597    | 10510   | RTA00002905F.d.17.1.P.Seq | F            | M00008001B:F05  | CH03MAH |
| 598    | 13539   | RTA00002898F.f.03.1.P.Seq | F            | M00004336A:A01  | CH01COH |
| 599    | 20149   | RTA00002917F.o.03.1.P.Seq | F            | M00032791D:F01  | CH08LNH |
| 600    | 12780   | RTA00002891F.e.06.1.P.Seq | F            | M00001686D:F06  | CH01COH |
| 601    | 182479  | RTA00002910F.i.18.1.P.Seq | F            | M00022972C:E05  | CH03MAH |
| 602    | 14016   | RTA00002923F.n.17.1.P.Seq | F            | M00039344C:A11  | CH09LNL |
| 603    | 76075   | RTA00002890F.h.23.1.P.Seq | F            | M00001620B:A03  | CH01COH |
| 604    | 9806    | RTA00002922F.i.05.1.P.Seq | F            | M00039078D:C10  | CH09LNL |
| 605    | 9086    | RTA00002889F.k.15.1.P.Seq | F            | M00001558A:E06  | CH01COH |
| 606    | 2619    | RTA00002907F.o.12.1.P.Seq | F            | M00022269C:A04  | CH03MAH |
| 607    | 17517   | RTA00002907F.h.06.1.P.Seq | F            | M00022185A:B03  | CH03MAH |
| 608    | 5089    | RTA00002915F.e.22.2.P.Seq | F            | M00028777B:G04  | CH08LNH |
| 609    | 6728    | RTA00002904F.b.13.1.P.Seq | F            | M00007178A:C02  | CH02COH |
| 610    | 41149   | RTA00002899F.g.20.1.P.Seq | F            | M00004603B:E02  | CH01COH |
| 611    | 35017   | RTA00002892F.f.03.2.P.Seq | F            | M00003812C:A03  | CH01COH |
| 612    | 7008    | RTA00002923F.a.08.1.P.Seq | F            | M00039165D:C04  | CH09LNL |
| 613    | 185545  | RTA00002912F.k.16.1.P.Seq | F            | M00027480C:E09  | CH04MAL |
| 614    | 17840   | RTA00002892F.p.15.2.P.Seq | F            | M00003854B:F07  | CH01COH |
| 615    | 185914  | RTA00002912F.i.24.1.P.Seq | F            | M00027467A:C07  | CH04MAL |
| 616    | 6862    | RTA00002903F.b.08.1.P.Seq | F            | M00006872D:B07  | CH02COH |
| 617    | 20120   | RTA00002888F.c.24.1.P.Seq | F            | M00001445B:F06  | CH01COH |
| 618    | 20120   | RTA00002888F.d.01.1.P.Seq | F            | M00001445B:F06  | CH01COH |
| 619    | 13879   | RTA00002923F.d.02.1.P.Seq | F            | M00039207A:F07  | CH09LNL |
| 620    | 9330    | RTA00002915F.g.16.1.P.Seq | F            | M00028786B:A04  | CH08LNH |
| 621    | 21572   | RTA00002921F.h.19.1.P.Seq | F            | M00033441A:B12  | CH09LNL |
| 622    | 2943    | RTA00002919F.h.22.1.P.Seq | F            | M00033144A:D02  | CH08LNH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 623    | 32154   | RTA00002905F.h.16.1.P.Seq | F            | M00007969D:C01 | CH03MAH   |
| 624    | 20875   | RTA00002901F.k.16.1.P.Seq | F            | M00005603B:H03 | CH02COH   |
| 625    | 186324  | RTA00002912F.d.17.1.P.Seq | F            | M00027274A:A09 | CH04MAL   |
| 626    | 10768   | RTA00002886F.e.01.1.P.Seq | F            | M00001346B:G11 | CH01COH   |
| 627    | 16711   | RTA00002935F.m.11.1.P.Seq | F            | M00055221C:H11 | CH17COHLV |
| 628    | 14688   | RTA00002925F.n.14.1.P.Seq | F            | M00040023B:B10 | CH09LNL   |
| 629    | 44419   | RTA00002907F.b.19.1.P.Seq | F            | M00022118A:E06 | CH03MAH   |
| 630    | 12614   | RTA00002896F.p.03.1.P.Seq | F            | M00004201D:C01 | CH01COH   |
| 631    | 21658   | RTA00002902F.c.23.1.P.Seq | F            | M00006576D:C02 | CH02COH   |
| 632    | 10150   | RTA00002901F.i.16.1.P.Seq | F            | M00005540A:F09 | CH02COH   |
| 633    | 185909  | RTA00002912F.c.20.1.P.Seq | F            | M00027262A:A07 | CH04MAL   |
| 634    | 14893   | RTA00002890F.f.08.1.P.Seq | F            | M00001607D:H09 | CH01COH   |
| 635    | 32125   | RTA00002903F.c.08.1.P.Seq | F            | M00006884D:A08 | CH02COH   |
| 636    | 11909   | RTA00002902F.a.11.1.P.Seq | F            | M00005766D:D12 | CH02COH   |
| 637    | 17237   | RTA00002901F.l.12.1.P.Seq | F            | M00005616B:F07 | CH02COH   |
| 638    | 11148   | RTA00002900F.j.18.1.P.Seq | F            | M00005346D:A03 | CH02COH   |
| 639    | 14837   | RTA00002925F.n.20.1.P.Seq | F            | M00040025A:B04 | CH09LNL   |
| 640    | 4343    | RTA00002897F.l.13.1.P.Seq | F            | M00004282B:D07 | CH01COH   |
| 641    | 18686   | RTA00002898F.i.16.1.P.Seq | F            | M00004366D:C11 | CH01COH   |
| 642    | 10090   | RTA00002892F.n.10.2.P.Seq | F            | M00003842D:H09 | CH01COH   |
| 643    | 612     | RTA00002889F.d.13.2.P.Seq | F            | M00001535B:E02 | CH01COH   |
| 644    | 10752   | RTA00002892F.n.06.2.P.Seq | F            | M00003842D:D11 | CH01COH   |
| 645    | 167203  | RTA00002914F.c.14.1.P.Seq | F            | M00028070A:H09 | CH08LNH   |
| 646    | 21269   | RTA00002901F.j.15.1.P.Seq | F            | M00005570A:B08 | CH02COH   |
| 647    | 186250  | RTA00002910F.a.21.1.P.Seq | F            | M00022797D:A06 | CH03MAH   |
| 648    | 24633   | RTA00002907F.i.19.2.P.Seq | F            | M00022208B:D03 | CH03MAH   |
| 649    | 12295   | RTA00002918F.c.02.1.P.Seq | F            | M00032836B:A07 | CH08LNH   |
| 650    | 7870    | RTA00002905F.b.22.1.P.Seq | F            | M00007973B:D11 | CH03MAH   |
| 651    | 12225   | RTA00002902F.d.08.1.P.Seq | F            | M00006585A:D07 | CH02COH   |
| 652    | 7775    | RTA00002892F.o.12.2.P.Seq | F            | M00003847A:H04 | CH01COH   |
| 653    | 14901   | RTA00002929F.f.21.1.P.Seq | F            | M00040349D:D07 | CH14EDT   |
| 654    | 6831    | RTA00002927F.b.21.1.P.Seq | F            | M00039483A:D10 | CH12EDT   |
| 655    | 10738   | RTA00002930F.b.08.1.P.Seq | F            | M00042724A:G06 | CH15CON   |
| 656    | 17986   | RTA00002932F.a.20.1.P.Seq | F            | M00042972C:F04 | CH18CON   |
| 657    | 23163   | RTA00002895F.h.03.1.P.Seq | F            | M00004085A:H01 | CH01COH   |
| 658    | 4838    | RTA00002923F.i.15.1.P.Seq | F            | M00039284D:H07 | CH09LNL   |
| 659    | 25386   | RTA00002905F.e.05.1.P.Seq | F            | M00008007B:E03 | CH03MAH   |
| 660    | 13217   | RTA00002887F.n.01.1.P.Seq | F            | M00001422B:D06 | CH01COH   |
| 661    | 30656   | RTA00002906F.l.03.1.P.Seq | F            | M00022032A:G05 | CH03MAH   |
| 662    | 7852    | RTA00002889F.e.14.1.P.Seq | F            | M00001538B:A07 | CH01COH   |
| 663    | 13217   | RTA00002887F.m.24.1.P.Seq | F            | M00001422B:D06 | CH01COH   |
| 664    | 15152   | RTA00002925F.f.24.1.P.Seq | F            | M00039873B:H04 | CH09LNL   |
| 665    | 24143   | RTA00002922F.o.18.1.P.Seq | F            | M00039143D:C10 | CH09LNL   |
| 666    | 23872   | RTA00002892F.i.13.1.P.Seq | F            | M00003823B:A06 | CH01COH   |
| 667    | 13940   | RTA00002906F.g.23.1.P.Seq | F            | M00021967D:H06 | CH03MAH   |
| 668    | 25759   | RTA00002907F.m.10.1.P.Seq | F            | M00022249D:C01 | CH03MAH   |
| 669    | 5761    | RTA00002924F.p.05.1.P.Seq | F            | M00039786D:A10 | CH09LNL   |
| 670    | 41703   | RTA00002901F.g.23.1.P.Seq | F            | M00005506D:E11 | CH02COH   |
| 671    | 7165    | RTA00002909F.i.06.1.P.Seq | F            | M00022648A:D08 | CH03MAH   |
| 672    | 41492   | RTA00002889F.m.18.1.P.Seq | F            | M00001565A:H05 | CH01COH   |
| 673    | 9331    | RTA00002906F.g.10.1.P.Seq | F            | M00021958B:E08 | CH03MAH   |
| 674    | 7961    | RTA00002887F.g.24.1.P.Seq | F            | M00001399B:B01 | CH01COH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|-------------|----------------|---------|
| 675    | 15367   | RTA00002893F.n.17.1.P.Seq | F           | M00003958C:H08 | CH01COH |
| 676    | 185628  | RTA00002912F.f.17.1.P.Seq | F           | M00027319C:C03 | CH04MAL |
| 677    | 7386    | RTA00002891F.l.14.1.P.Seq | F           | M00003768D:D08 | CH01COH |
| 678    | 67391   | RTA00002893F.p.07.1.P.Seq | F           | M00003968C:G03 | CH01COH |
| 679    | 46380   | RTA00002906F.f.10.1.P.Seq | F           | M00021933B:F02 | CH03MAH |
| 680    | 14265   | RTA00002892F.e.05.2.P.Seq | F           | M00003808A:F11 | CH01COH |
| 681    | 186478  | RTA00002912F.f.07.1.P.Seq | F           | M00027313C:E01 | CH04MAL |
| 682    | 8192    | RTA00002916F.m.07.1.P.Seq | F           | M00032634B:D09 | CH08LNH |
| 683    | 13776   | RTA00002925F.l.10.1.P.Seq | F           | M00039976C:F11 | CH09LNL |
| 684    | 11796   | RTA00002912F.e.02.1.P.Seq | F           | M00027291A:G08 | CH04MAL |
| 685    | 10827   | RTA00002919F.i.10.1.P.Seq | F           | M00033147C:B08 | CH08LNH |
| 686    | 1482    | RTA00002925F.l.12.1.P.Seq | F           | M00039977B:D12 | CH09LNL |
| 687    | 30300   | RTA00002906F.f.16.1.P.Seq | F           | M00021941A:D09 | CH03MAH |
| 688    | 10454   | RTA00002890F.i.15.1.P.Seq | F           | M00001623D:A10 | CH01COH |
| 689    | 16649   | RTA00002907F.l.01.1.P.Seq | F           | M00022229D:E01 | CH03MAH |
| 690    | 7026    | RTA00002887F.b.10.1.P.Seq | F           | M00001387B:A11 | CH01COH |
| 691    | 5691    | RTA00002895F.n.13.1.P.Seq | F           | M00004114C:D11 | CH01COH |
| 692    | 13797   | RTA00002918F.i.21.1.P.Seq | F           | M00032918D:B04 | CH08LNH |
| 693    | 5187    | RTA00002923F.n.03.1.P.Seq | F           | M00039338B:F07 | CH09LNL |
| 694    | 186115  | RTA00002912F.i.01.1.P.Seq | F           | M00027376C:A02 | CH04MAL |
| 695    | 4826    | RTA00002917F.g.24.1.P.Seq | F           | M00032729A:F10 | CH08LNH |
| 696    | 6733    | RTA00002917F.m.11.1.P.Seq | F           | M00032774C:C04 | CH08LNH |
| 697    | 7604    | RTA00002923F.j.05.1.P.Seq | F           | M00039291D:F02 | CH09LNL |
| 698    | 46459   | RTA00002905F.f.01.1.P.Seq | F           | M00008020D:F02 | CH03MAH |
| 699    | 23385   | RTA00002889F.i.23.1.P.Seq | F           | M00001551D:D01 | CH01COH |
| 700    | 7516    | RTA00002891F.h.11.1.P.Seq | F           | M00003749C:C08 | CH01COH |
| 701    | 45048   | RTA00002906F.b.04.1.P.Seq | F           | M00021855D:F10 | CH03MAH |
| 702    | 14845   | RTA00002903F.o.02.1.P.Seq | F           | M00007103C:C12 | CH02COH |
| 703    | 16479   | RTA00002887F.i.15.1.P.Seq | F           | M00001403D:C12 | CH01COH |
| 704    | 186729  | RTA00002911F.d.19.2.P.Seq | F           | M00026850B:C09 | CH04MAL |
| 705    | 33658   | RTA00002886F.j.07.1.P.Seq | F           | M00001361B:A12 | CH01COH |
| 706    | 186755  | RTA00002912F.i.18.1.P.Seq | F           | M00027400D:H02 | CH04MAL |
| 707    | 4262    | RTA00002897F.a.04.1.P.Seq | F           | M00004208A:D08 | CH01COH |
| 708    | 14039   | RTA00002897F.k.01.1.P.Seq | F           | M00004276C:A08 | CH01COH |
| 709    | 11948   | RTA00002895F.n.24.1.P.Seq | F           | M00004118C:D12 | CH01COH |
| 710    | 14865   | RTA00002908F.l.14.1.P.Seq | F           | M00022481B:A04 | CH03MAH |
| 711    | 10779   | RTA00002915F.i.01.1.P.Seq | F           | M00031370B:C01 | CH08LNH |
| 712    | 7503    | RTA00002902F.k.16.1.P.Seq | F           | M00006738A:F12 | CH02COH |
| 713    | 48130   | RTA00002902F.e.04.1.P.Seq | F           | M00006595B:C10 | CH02COH |
| 714    | 7858    | RTA00002907F.m.12.1.P.Seq | F           | M00022250A:B04 | CH03MAH |
| 715    | 4682    | RTA00002924F.n.02.1.P.Seq | F           | M00039710B:E01 | CH09LNL |
| 716    | 20650   | RTA00002888F.p.10.1.P.Seq | F           | M00001503B:H10 | CH01COH |
| 717    | 25320   | RTA00002910F.e.19.1.P.Seq | F           | M00022857B:A09 | CH03MAH |
| 718    | 4924    | RTA00002930F.g.01.2.P.Seq | F           | M00055805A:H02 | CH15CON |
| 719    | 21170   | RTA00002900F.l.13.1.P.Seq | F           | M00005365A:F05 | CH02COH |
| 720    | 9258    | RTA00002890F.h.17.1.P.Seq | F           | M00001618C:D01 | CH01COH |
| 721    | 14039   | RTA00002897F.j.24.1.P.Seq | F           | M00004276C:A08 | CH01COH |
| 722    | 3483    | RTA00002899F.b.07.1.P.Seq | F           | M00004430A:A05 | CH01COH |
| 723    | 3877    | RTA00002897F.k.10.1.P.Seq | F           | M00004278A:G06 | CH01COH |
| 724    | 7483    | RTA00002923F.f.19.1.P.Seq | F           | M00039246B:A08 | CH09LNL |
| 725    | 99750   | RTA00002900F.l.17.1.P.Seq | F           | M00005366D:F08 | CH02COH |
| 726    | 46459   | RTA00002905F.e.24.1.P.Seq | F           | M00008020D:F02 | CH03MAH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|-------------|----------------|---------|
| 727    | 3591    | RTA00002886F.a.09.1.P.Seq | F           | M00001378C:E10 | CH01COH |
| 728    | 11277   | RTA00002923F.g.11.1.P.Seq | F           | M00039251D:B08 | CH09LNL |
| 729    | 10292   | RTA00002898F.n.11.1.P.Seq | F           | M00004393C:D06 | CH01COH |
| 730    | 23211   | RTA00002922F.k.13.1.P.Seq | F           | M00039105D:A08 | CH09LNL |
| 731    | 185698  | RTA00002911F.d.03.2.P.Seq | F           | M00026836B:H03 | CH04MAL |
| 732    | 24702   | RTA00002898F.i.17.1.P.Seq | F           | M00004360C:D09 | CH01COH |
| 733    | 12595   | RTA00002904F.c.06.1.P.Seq | F           | M00007197B:B05 | CH02COH |
| 734    | 177444  | RTA00002910F.o.06.1.P.Seq | F           | M00023097D:B08 | CH03MAH |
| 735    | 38147   | RTA00002886F.o.11.1.P.Seq | F           | M00001379A:F09 | CH01COH |
| 736    | 17909   | RTA00002908F.d.09.1.P.Seq | F           | M00022386D:F10 | CH03MAH |
| 737    | 13399   | RTA00002900F.n.07.1.P.Seq | F           | M00005385A:B12 | CH02COH |
| 738    | 17720   | RTA00002905F.g.22.1.P.Seq | F           | M00008065D:A07 | CH03MAH |
| 739    | 45974   | RTA00002909F.l.08.1.P.Seq | F           | M00022681D:E10 | CH03MAH |
| 740    | 10779   | RTA00002915F.h.24.2.P.Seq | F           | M00031370B:C01 | CH08LNL |
| 741    | 195     | RTA00002914F.a.14.1.P.Seq | F           | M00028055B:G07 | CH08LNL |
| 742    | 1712    | RTA00002915F.j.07.1.P.Seq | F           | M00031484A:D03 | CH08LNL |
| 743    | 185726  | RTA00002912F.a.21.1.P.Seq | F           | M00027215B:B12 | CH04MAL |
| 744    | 150298  | RTA00002907F.d.20.1.P.Seq | F           | M00022140D:A07 | CH03MAH |
| 745    | 358     | RTA00002898F.i.02.1.P.Seq | F           | M00004358B:G02 | CH01COH |
| 746    | 42920   | RTA00002900F.i.16.1.P.Seq | F           | M00005309B:A11 | CH02COH |
| 747    | 25681   | RTA00002894F.k.18.1.P.Seq | F           | M00004038A:A04 | CH01COH |
| 748    | 18005   | RTA00002903F.c.18.1.P.Seq | F           | M00006890C:F10 | CH02COH |
| 749    | 16143   | RTA00002892F.m.04.2.P.Seq | F           | M00003839C:H10 | CH01COH |
| 750    | 9306    | RTA00002902F.d.09.1.P.Seq | F           | M00006585A:F09 | CH02COH |
| 751    | 32293   | RTA00002901F.i.13.1.P.Seq | F           | M00005535B:B01 | CH02COH |
| 752    | 8913    | RTA00002901F.j.07.1.P.Seq | F           | M00005557D:H10 | CH02COH |
| 753    | 185819  | RTA00002912F.a.20.1.P.Seq | F           | M00027215A:F06 | CH04MAL |
| 754    | 10559   | RTA00002898F.o.12.1.P.Seq | F           | M00004406A:G09 | CH01COH |
| 755    | 8740    | RTA00002923F.o.11.1.P.Seq | F           | M00039383A:H07 | CH09LNL |
| 756    | 160257  | RTA00002907F.l.12.2.P.Seq | F           | M00022237C:E04 | CH03MAH |
| 757    | 6078    | RTA00002930F.c.11.1.P.Seq | F           | M00055433D:G03 | CH15CON |
| 758    | 12543   | RTA00002927F.b.14.1.P.Seq | F           | M00039377B:E05 | CH12EDT |
| 759    | 9686    | RTA00002930F.f.19.1.P.Seq | F           | M00055794A:E10 | CH15CON |
| 760    | 3369    | RTA00002930F.b.12.1.P.Seq | F           | M00042732B:H06 | CH15CON |
| 761    | 6891    | RTA00002895F.i.03.1.P.Seq | F           | M00004087C:E02 | CH01COH |
| 762    | 13666   | RTA00002892F.i.05.1.P.Seq | F           | M00003822C:A09 | CH01COH |
| 763    | 6925    | RTA00002930F.k.24.1.P.Seq | F           | M00056458C:E01 | CH15CON |
| 764    | 11351   | RTA00002901F.g.15.1.P.Seq | F           | M00005504D:F06 | CH02COH |
| 765    | 11497   | RTA00002889F.a.21.1.P.Seq | F           | M00001512D:F08 | CH01COH |
| 766    | 1596    | RTA00002922F.m.18.1.P.Seq | F           | M00039125D:H12 | CH09LNL |
| 767    | 186519  | RTA00002924F.a.22.1.P.Seq | F           | M00039411D:D09 | CH09LNL |
| 768    | 24429   | RTA00002903F.j.04.1.P.Seq | F           | M00006989B:G05 | CH02COH |
| 769    | 33795   | RTA00002902F.k.18.1.P.Seq | F           | M00006739B:A04 | CH02COH |
| 770    | 24267   | RTA00002889F.l.17.1.P.Seq | F           | M00001561D:H04 | CH01COH |
| 771    | 12536   | RTA00002891F.j.20.1.P.Seq | F           | M00003760C:G10 | CH01COH |
| 772    | 22627   | RTA00002887F.k.07.1.P.Seq | F           | M00001410A:G10 | CH01COH |
| 773    | 24430   | RTA00002901F.h.20.1.P.Seq | F           | M00005520B:E01 | CH02COH |
| 774    | 16151   | RTA00002897F.l.22.1.P.Seq | F           | M00004284A:F08 | CH01COH |
| 775    | 6148    | RTA00002890F.i.16.1.P.Seq | F           | M00001623D:E12 | CH01COH |
| 776    | 106064  | RTA00002908F.l.19.1.P.Seq | F           | M00022485B:E07 | CH03MAH |
| 777    | 9573    | RTA00002893F.p.13.1.P.Seq | F           | M00003970D:H07 | CH01COH |
| 778    | 19542   | RTA00002902F.l.20.1.P.Seq | F           | M00006756B:G06 | CH02COH |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 779    | 16672   | RTA00002889F.h.21.1.P.Seq | F           | M00001528C:C03 | CH01COH   |
| 780    | 8573    | RTA00002891F.p.07.1.P.Seq | F           | M00003785D:F07 | CH01COH   |
| 781    | 15746   | RTA00002896F.h.10.1.P.Seq | F           | M00004163C:A03 | CH01COH   |
| 782    | 4500    | RTA00002887F.b.08.1.P.Seq | F           | M00001387A:C12 | CH01COH   |
| 783    | 16003   | RTA00002910F.c.08.1.P.Seq | F           | M00022820A:F07 | CH03MAH   |
| 784    | 18723   | RTA00002916F.g.18.1.P.Seq | F           | M00032580D:A09 | CH08LNL   |
| 785    | 4270    | RTA00002922F.b.01.1.P.Seq | F           | M00038616C:C09 | CH09LNL   |
| 786    | 30095   | RTA00002907F.i.20.1.P.Seq | F           | M00022208C:E04 | CH03MAH   |
| 787    | 42916   | RTA00002924F.c.08.1.P.Seq | F           | M00039433B:D06 | CH09LNL   |
| 788    | 13652   | RTA00002902F.j.09.1.P.Seq | F           | M00006714C:D06 | CH02COH   |
| 789    | 6972    | RTA00002902F.i.06.1.P.Seq | F           | M00006712C:H01 | CH02COH   |
| 790    | 4519    | RTA00002910F.i.06.1.P.Seq | F           | M00022947B:D02 | CH03MAH   |
| 791    | 13106   | RTA00002928F.f.09.1.P.Seq | F           | M00040224C:F06 | CH13EDT   |
| 792    | 98186   | RTA00002909F.m.08.1.P.Seq | F           | M00022696B:C11 | CH03MAH   |
| 793    | 3167    | RTA00002898F.g.09.1.P.Seq | F           | M00004344D:C12 | CH01COH   |
| 794    | 3272    | RTA00002897F.a.18.1.P.Seq | F           | M00004212D:C03 | CH01COH   |
| 795    | 14446   | RTA00002899F.d.05.1.P.Seq | F           | M00004462D:D12 | CH01COH   |
| 796    | 17865   | RTA00002918F.a.13.1.P.Seq | F           | M00032825B:F08 | CH08LNL   |
| 797    | 5834    | RTA00002898F.h.12.1.P.Seq | F           | M00004352A:D08 | CH01COH   |
| 798    | 14533   | RTA00002896F.k.24.1.P.Seq | F           | M00004179C:B06 | CH01COH   |
| 799    | 15222   | RTA00002900F.j.05.1.P.Seq | F           | M00005332A:C06 | CH02COH   |
| 800    | 22594   | RTA00002898F.h.21.1.P.Seq | F           | M00004357B:B06 | CH01COH   |
| 801    | 9204    | RTA00002890F.h.20.1.P.Seq | F           | M00001619C:H09 | CH01COH   |
| 802    | 186464  | RTA00002911F.d.09.2.P.Seq | F           | M00026842D:C02 | CH04MAL   |
| 803    | 5441    | RTA00002900F.a.11.1.P.Seq | F           | M00004824D:H05 | CH02COH   |
| 804    | 32544   | RTA00002893F.l.21.1.P.Seq | F           | M00003935B:B01 | CH01COH   |
| 805    | 15351   | RTA00002915F.j.15.1.P.Seq | F           | M00032471D:A05 | CH08LNL   |
| 806    | 13129   | RTA00002898F.a.12.1.P.Seq | F           | M00004310B:E02 | CH01COH   |
| 807    | 186376  | RTA00002912F.k.21.1.P.Seq | F           | M00027485C:F07 | CH04MAL   |
| 808    | 17816   | RTA00002901F.o.04.1.P.Seq | F           | M00005674C:F04 | CH02COH   |
| 809    | 8434    | RTA00002923F.l.22.1.P.Seq | F           | M00039326C:B08 | CH09LNL   |
| 810    | 22146   | RTA00002922F.i.08.1.P.Seq | F           | M00039067B:F07 | CH09LNL   |
| 811    | 31912   | RTA00002904F.a.14.1.P.Seq | F           | M00007154A:E06 | CH02COH   |
| 812    | 1487    | RTA00002925F.n.03.1.P.Seq | F           | M00040016C:E07 | CH09LNL   |
| 813    | 24777   | RTA00002900F.n.02.1.P.Seq | F           | M00005380B:H10 | CH02COH   |
| 814    | 144483  | RTA00002902F.d.01.1.P.Seq | F           | M00006577A:H10 | CH02COH   |
| 815    | 6546    | RTA00002935F.p.16.1.P.Seq | F           | M00055425C:A04 | CH17COHLV |
| 816    | 5984    | RTA00002935F.p.09.1.P.Seq | F           | M00055420A:E06 | CH17COHLV |
| 817    | 24441   | RTA00002900F.a.22.1.P.Seq | F           | M00004832D:G04 | CH02COH   |
| 818    | 20889   | RTA00002935F.h.09.1.P.Seq | F           | M00054807D:C11 | CH17COHLV |
| 819    | 127721  | RTA00002915F.c.18.1.P.Seq | F           | M00028763A:G11 | CH08LNL   |
| 820    | 20684   | RTA00002900F.c.03.1.P.Seq | F           | M00004843A:G12 | CH02COH   |
| 821    | 30095   | RTA00002907F.i.20.2.P.Seq | F           | M00022208C:E04 | CH03MAH   |
| 822    | 6763    | RTA00002892F.o.01.2.P.Seq | F           | M00003845D:G03 | CH01COH   |
| 823    | 6763    | RTA00002892F.n.24.2.P.Seq | F           | M00003845D:G03 | CH01COH   |
| 824    | 48725   | RTA00002907F.l.22.2.P.Seq | F           | M00022240B:C12 | CH03MAH   |
| 825    | 21260   | RTA00002935F.c.22.1.P.Seq | F           | M00054499A:C08 | CH17COHLV |
| 826    | 42572   | RTA00002930F.c.21.1.P.Seq | F           | M00055454A:D02 | CH15CON   |
| 827    | 3441    | RTA00002935F.i.13.1.P.Seq | F           | M00054890C:D05 | CH17COHLV |
| 828    | 21419   | RTA00002930F.b.13.1.P.Seq | F           | M00042734A:F05 | CH15CON   |
| 829    | 8004    | RTA00002910F.b.08.1.P.Seq | F           | M00022805B:A10 | CH03MAH   |
| 830    | 185870  | RTA00002912F.c.06.1.P.Seq | F           | M00027247C:D02 | CH04MAL   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 831    | 24580   | RTA00002930F.d.01.1.P.Seq | F            | M00055466A:F06 | CH15CON   |
| 832    | 5153    | RTA00002930F.b.16.1.P.Seq | F            | M00042743D:G10 | CH15CON   |
| 833    | 8653    | RTA00002895F.f.17.1.P.Seq | F            | M00004080C:C04 | CH01COH   |
| 834    | 23799   | RTA00002924F.l.23.1.P.Seq | F            | M00039698C:B03 | CH09LNL   |
| 835    | 11012   | RTA00002930F.j.09.1.P.Seq | F            | M00056215D:F02 | CH15CON   |
| 836    | 46592   | RTA00002900F.b.19.1.P.Seq | F            | M00004839B:C12 | CH02COH   |
| 837    | 6650    | RTA00002908F.m.12.1.P.Seq | F            | M00022491D:A10 | CH03MAH   |
| 838    | 16618   | RTA00002889F.n.18.1.P.Seq | F            | M00001568C:A03 | CH01COH   |
| 839    | 18274   | RTA00002889F.g.05.1.P.Seq | F            | M00001543C:A08 | CH01COH   |
| 840    | 20694   | RTA00002908F.h.08.1.P.Seq | F            | M00022442B:G03 | CH03MAH   |
| 841    | 9493    | RTA00002909F.m.11.1.P.Seq | F            | M00022698C:D10 | CH03MAH   |
| 842    | 6132    | RTA00002897F.c.04.1.P.Seq | F            | M00004220D:C11 | CH01COH   |
| 843    | 186259  | RTA00002912F.m.13.1.P.Seq | F            | M00027527B:C05 | CH04MAL   |
| 844    | 3769    | RTA00002916F.g.22.1.P.Seq | F            | M00032581B:A09 | CH08LNL   |
| 845    | 36584   | RTA00002935F.f.12.1.P.Seq | F            | M00054683D:G11 | CH17COHLV |
| 846    | 38077   | RTA00002890F.e.06.1.P.Seq | F            | M00001605B:B05 | CH01COH   |
| 847    | 3927    | RTA00002935F.a.12.1.P.Seq | F            | M00042516B:D01 | CH17COHLV |
| 848    | 4275    | RTA00002914F.b.16.1.P.Seq | F            | M00028063C:H01 | CH08LNL   |
| 849    | 12554   | RTA00002921F.a.23.1.P.Seq | F            | M00033302A:E11 | CH09LNL   |
| 850    | 13761   | RTA00002901F.f.22.1.P.Seq | F            | M00005489B:C08 | CH02COH   |
| 851    | 19059   | RTA00002897F.e.22.1.P.Seq | F            | M00004237C:D10 | CH01COH   |
| 852    | 22944   | RTA00002935F.b.17.1.P.Seq | F            | M00043355A:D07 | CH17COHLV |
| 853    | 2189    | RTA00002925F.i.06.1.P.Seq | F            | M00039921A:B10 | CH09LNL   |
| 854    | 19153   | RTA00002892f.H.04.2.P.Seq | F            | M00003819B:B01 | CH01COH   |
| 855    | 1833    | RTA00002890F.e.13.1.P.Seq | F            | M00001606B:A10 | CH01COH   |
| 856    | 18447   | RTA00002935F.d.23.1.P.Seq | F            | M00054569A:B07 | CH17COHLV |
| 857    | 2461    | RTA00002922F.b.08.1.P.Seq | F            | M00038619B:F09 | CH09LNL   |
| 858    | 15917   | RTA00002896F.i.06.1.P.Seq | F            | M00004172C:A08 | CH01COH   |
| 859    | 9379    | RTA00002935F.a.15.1.P.Seq | F            | M00043299A:B10 | CH17COHLV |
| 860    | 5511    | RTA00002931F.b.06.1.P.Seq | F            | M00042796A:A10 | CH16COP   |
| 861    | 10540   | RTA00002891F.k.16.1.P.Seq | F            | M00003764B:H11 | CH01COH   |
| 862    | 12117   | RTA00002899F.a.09.1.P.Seq | F            | M00004419A:G02 | CH01COH   |
| 863    | 8777    | RTA00002919F.a.23.1.P.Seq | F            | M00033028D:C10 | CH08LNL   |
| 864    | 23972   | RTA00002900F.o.18.1.P.Seq | F            | M00005403C:A01 | CH02COH   |
| 865    | 17005   | RTA00002896F.m.10.1.P.Seq | F            | M00004187B:C02 | CH01COH   |
| 866    | 1085    | RTA00002924F.l.20.1.P.Seq | F            | M00039694C:H01 | CH09LNL   |
| 867    | 4270    | RTA00002922F.a.24.1.P.Seq | F            | M00038616C:C09 | CH09LNL   |
| 868    | 4609    | RTA00002935F.e.15.1.P.Seq | F            | M00054599D:B03 | CH17COHLV |
| 869    | 6889    | RTA00002919F.c.07.1.P.Seq | F            | M00033037B:F04 | CH08LNL   |
| 870    | 15228   | RTA00002919F.e.06.1.P.Seq | F            | M00033055D:D02 | CH08LNL   |
| 871    | 20971   | RTA00002904F.a.22.1.P.Seq | F            | M00007158D:D03 | CH02COH   |
| 872    | 5174    | RTA00002935F.a.23.1.P.Seq | F            | M00043313D:E09 | CH17COHLV |
| 873    | 15236   | RTA00002928F.e.16.1.P.Seq | F            | M00040198A:F12 | CH13EDT   |
| 874    | 9223    | RTA00002896F.b.15.1.P.Seq | F            | M00004141A:D01 | CH01COH   |
| 875    | 24591   | RTA00002923F.g.10.1.P.Seq | F            | M00039251C:H12 | CH09LNL   |
| 876    | 36306   | RTA00002888F.l.11.1.P.Seq | F            | M00001485C:F06 | CH01COH   |
| 877    | 3309    | RTA00002893F.i.21.1.P.Seq | F            | M00003916A:E04 | CH01COH   |
| 878    | 186712  | RTA00002911F.c.11.2.P.Seq | F            | M00026809A:H08 | CH04MAL   |
| 879    | 9090    | RTA00002891F.g.23.1.P.Seq | F            | M00003746C:E11 | CH01COH   |
| 880    | 11510   | RTA00002888F.i.07.1.P.Seq | F            | M00001467C:D04 | CH01COH   |
| 881    | 9784    | RTA00002889F.i.15.1.P.Seq | F            | M00001554C:G10 | CH01COH   |
| 882    | 25618   | RTA00002930F.a.11.1.P.Seq | F            | M00042554A:D01 | CH15CON   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 883    | 12493   | RTA00002928F.i.11.1.P.Seq | F           | M00040289D:C06 | CH13EDT   |
| 884    | 24361   | RTA00002933F.b.08.1.P.Seq | F           | M00043134A:F05 | CH19COP   |
| 885    | 12449   | RTA00002930F.d.02.1.P.Seq | F           | M00055468A:A08 | CH15CON   |
| 886    | 17894   | RTA00002929F.a.04.1.P.Seq | F           | M00039747B:B06 | CH14EDT   |
| 887    | 13204   | RTA00002930F.f.09.1.P.Seq | F           | M00055745B:A08 | CH15CON   |
| 888    | 32119   | RTA00002930F.c.19.1.P.Seq | F           | M00055448B:E05 | CH15CON   |
| 889    | 5909    | RTA00002935F.i.23.1.P.Seq | F           | M00054931D:E10 | CH17COHLV |
| 890    | 24453   | RTA00002927F.d.15.1.P.Seq | F           | M00039526A:A08 | CH12EDT   |
| 891    | 46982   | RTA00002935F.k.22.1.P.Seq | F           | M00055093B:A03 | CH17COHLV |
| 892    | 43888   | RTA00002932F.b.23.1.P.Seq | F           | M00043070A:C03 | CH18CON   |
| 893    | 24580   | RTA00002930F.c.24.1.P.Seq | F           | M00055466A:F06 | CH15CON   |
| 894    | 186495  | RTA00002927F.a.21.1.P.Seq | F           | M00039364D:E05 | CH12EDT   |
| 895    | 12420   | RTA00002932F.b.21.1.P.Seq | F           | M00043063C:H05 | CH18CON   |
| 896    | 3833    | RTA00002916F.e.14.1.P.Seq | F           | M00032562C:F01 | CH08LNH   |
| 897    | 10438   | RTA00002930F.i.13.1.P.Seq | F           | M00056230D:307 | CH15CON   |
| 898    | 12367   | RTA00002922F.n.10.1.P.Seq | F           | M00039133B:D06 | CH09LNL   |
| 899    | 5012    | RTA00002930F.k.16.1.P.Seq | F           | M00056342A:C03 | CH15CON   |
| 900    | 6458    | RTA00002929F.c.21.1.P.Seq | F           | M00040291A:G10 | CH14EDT   |
| 901    | 16507   | RTA00002929F.d.08.1.P.Seq | F           | M00040298B:B09 | CH14EDT   |
| 902    | 13914   | RTA00002922F.h.18.1.P.Seq | F           | M00039063C:H09 | CH09LNL   |
| 903    | 11590   | RTA00002930F.k.18.1.P.Seq | F           | M00056436C:F01 | CH15CON   |
| 904    | 15380   | RTA00002928F.i.06.1.P.Seq | F           | M00040287A:C11 | CH13EDT   |
| 905    | 10190   | RTA00002895F.k.07.1.P.Seq | F           | M00004096D:F02 | CH01COH   |
| 906    | 12593   | RTA00002934F.a.12.1.P.Seq | F           | M00043485C:C03 | CH20COHLV |
| 907    | 112813  | RTA00002905F.o.14.1.P.Seq | F           | M00021677A:D09 | CH03MAH   |
| 908    | 15929   | RTA00002930F.i.18.1.P.Seq | F           | M00056244C:H05 | CH15CON   |
| 909    | 16670   | RTA00002935F.o.16.1.P.Seq | F           | M00055387C:C12 | CH17COHLV |
| 910    | 10924   | RTA00002907F.k.12.2.P.Seq | F           | M00022224A:C07 | CH03MAH   |
| 911    | 6233    | RTA00002896F.b.17.1.P.Seq | F           | M00004141B:B01 | CH01COH   |
| 912    | 14777   | RTA00002897F.k.09.1.P.Seq | F           | M00004277D:B02 | CH01COH   |
| 913    | 12797   | RTA00002935F.h.01.1.P.Seq | F           | M00054781D:A11 | CH17COHLV |
| 914    | 186041  | RTA00002912F.c.01.1.P.Seq | F           | M00027244C:B06 | CH04MAL   |
| 915    | 8182    | RTA00002931F.a.22.1.P.Seq | F           | M00042766C:D05 | CH16COP   |
| 916    | 23088   | RTA00002888F.p.20.1.P.Seq | F           | M00001506B:D11 | CH01COH   |
| 917    | 24298   | RTA00002935F.p.18.1.P.Seq | F           | M00055473C:F02 | CH17COHLV |
| 918    | 40621   | RTA00002896F.k.12.1.P.Seq | F           | M00004176C:A09 | CH01COH   |
| 919    | 7124    | RTA00002935F.b.07.1.P.Seq | F           | M00043328C:E04 | CH17COHLV |
| 920    | 21107   | RTA00002901F.i.02.1.P.Seq | F           | M00005524C:H04 | CH02COH   |
| 921    | 10807   | RTA00002928F.c.15.1.P.Seq | F           | M00040162A:E02 | CH13EDT   |
| 922    | 12162   | RTA00002915F.j.23.1.P.Seq | F           | M00032475A:A06 | CH08LNH   |
| 923    | 14747   | RTA00002931F.a.18.1.P.Seq | F           | M00042512D:D10 | CH16COP   |
| 924    | 6824    | RTA00002931F.b.23.1.P.Seq | F           | M00042857C:E01 | CH16COP   |
| 925    | 39115   | RTA00002932F.a.17.1.P.Seq | F           | M00042967D:C01 | CH18CON   |
| 926    | 9484    | RTA00002934F.a.13.1.P.Seq | F           | M00043490C:F02 | CH20COHLV |
| 927    | 77981   | RTA00002890F.j.21.1.P.Seq | F           | M00001633D:C11 | CH01COH   |
| 928    | 16061   | RTA00002932F.b.24.1.P.Seq | F           | M00043113C:G09 | CH18CON   |
| 929    | 4834    | RTA00002930F.d.12.1.P.Seq | F           | M00055527B:E01 | CH15CON   |
| 930    | 9427    | RTA00002935F.e.21.1.P.Seq | F           | M00054623C:F05 | CH17COHLV |
| 931    | 167736  | RTA00002935F.l.11.1.P.Seq | F           | M00055117A:E02 | CH17COHLV |
| 932    | 16524   | RTA00002935F.b.19.1.P.Seq | F           | M00043358C:A02 | CH17COHLV |
| 933    | 23496   | RTA00002932F.a.21.1.P.Seq | F           | M00042976D:C01 | CH18CON   |
| 934    | 163647  | RTA00002901F.m.17.1.P.Seq | F           | M00005634A:F07 | CH02COH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 935    | 14239   | RTA00002931F.h.07.1.P.Seq | F            | M00042801C:D01 | CH16COP   |
| 936    | 25574   | RTA00002886F.g.20.1.P.Seq | F            | M00001353C:A05 | CH01COH   |
| 937    | 2737    | RTA00002932F.a.08.1.P.Seq | F            | M00042588C:E02 | CH18CON   |
| 938    | 6925    | RTA00002930F.l.01.1.P.Seq | F            | M00056458C:E01 | CH15CON   |
| 939    | 21106   | RTA00002925F.p.16.1.P.Seq | F            | M00040045B:H07 | CH09LNL   |
| 940    | 28134   | RTA00002917F.k.23.1.P.Seq | F            | M00032765A:C05 | CH08LNL   |
| 941    | 186496  | RTA00002917F.l.16.1.P.Seq | F            | M00032770C:G11 | CH08LNL   |
| 942    | 21625   | RTA00002893F.i.18.1.P.Seq | F            | M00003915C:G08 | CH01COH   |
| 943    | 12537   | RTA00002930F.g.23.1.P.Seq | F            | M00055919B:C10 | CH15CON   |
| 944    | 15577   | RTA00002902F.f.15.1.P.Seq | F            | M00006636B:E04 | CH02COH   |
| 945    | 6106    | RTA00002935F.f.11.1.P.Seq | F            | M00054682B:H02 | CH17COHLV |
| 946    | 17136   | RTA00002935F.h.14.1.P.Seq | F            | M00054818B:F10 | CH17COHLV |
| 947    | 2582    | RTA00002921F.f.14.1.P.Seq | F            | M00033413A:A08 | CH09LNL   |
| 948    | 16638   | RTA00002899F.a.15.1.P.Seq | F            | M00004420D:E05 | CH01COH   |
| 949    | 8869    | RTA00002929F.a.05.1.P.Seq | F            | M00039748C:G09 | CH14EDT   |
| 950    | 14426   | RTA00002914F.c.12.1.P.Seq | F            | M00028069D:H02 | CH08LNL   |
| 951    | 11994   | RTA00002890F.h.14.1.P.Seq | F            | M00001617C:F10 | CH01COH   |
| 952    | 186664  | RTA00002932F.a.05.1.P.Seq | F            | M00042585D:D03 | CH18CON   |
| 953    | 162235  | RTA00002907F.j.06.2.P.Seq | F            | M00022212D:G02 | CH03MAH   |
| 954    | 2127    | RTA00002912F.o.14.1.P.Seq | F            | M00027605B:D09 | CH04MAL   |
| 955    | 41014   | RTA00002901F.n.04.1.P.Seq | F            | M00005641B:E09 | CH02COH   |
| 956    | 17636   | RTA00002933F.c.19.1.P.Seq | F            | M00043222C:B06 | CH19COP   |
| 957    | 2328    | RTA00002935F.e.05.1.P.Seq | F            | M00054579A:C02 | CH17COHLV |
| 958    | 15414   | RTA00002935F.p.13.1.P.Seq | F            | M00055423C:H10 | CH17COHLV |
| 959    | 11948   | RTA00002895F.o.01.1.P.Seq | F            | M00004118C:D12 | CH01COH   |
| 960    | 24759   | RTA00002903F.n.05.1.P.Seq | F            | M00007082D:E05 | CH02COH   |
| 961    | 15152   | RTA00002925F.g.01.1.P.Seq | F            | M00039873B:H04 | CH09LNL   |
| 962    | 14917   | RTA00002922F.b.02.1.P.Seq | F            | M00038616D:B07 | CH09LNL   |
| 963    | 12941   | RTA00002889F.c.15.1.P.Seq | F            | M00001532A:G08 | CH01COH   |
| 964    | 29676   | RTA00002931F.b.03.1.P.Seq | F            | M00042788A:F04 | CH16COP   |
| 965    | 17789   | RTA00002891F.a.21.1.P.Seq | F            | M00001671A:H10 | CH01COH   |
| 966    | 45097   | RTA00002928F.g.06.1.P.Seq | F            | M00040247D:D02 | CH13EDT   |
| 967    | 18407   | RTA00002909F.b.11.1.P.Seq | F            | M00022546B:E05 | CH03MAH   |
| 968    | 22309   | RTA00002900F.n.19.1.P.Seq | F            | M00005392A:G06 | CH02COH   |
| 969    | 109382  | RTA00002907F.k.13.1.P.Seq | F            | M00022224A:G07 | CH03MAH   |
| 970    | 92273   | RTA00002909F.i.17.1.P.Seq | F            | M00022662D:H03 | CH03MAH   |
| 971    | 8403    | RTA00002915F.j.22.1.P.Seq | F            | M00032474A:G03 | CH08LNL   |
| 972    | 7763    | RTA00002928F.h.10.1.P.Seq | F            | M00040267D:A12 | CH13EDT   |
| 973    | 13470   | RTA00002930F.k.09.1.P.Seq | F            | M00056304A:H05 | CH15CON   |
| 974    | 1484    | RTA00002921F.k.10.1.P.Seq | F            | M00033556D:C10 | CH09LNL   |
| 975    | 10345   | RTA00002892F.o.19.2.P.Seq | F            | M00003848C:G09 | CH01COH   |
| 976    | 17242   | RTA00002931F.a.05.1.P.Seq | F            | M00042433A:E11 | CH16COP   |
| 977    | 171180  | RTA00002909F.f.24.1.P.Seq | F            | M00022618B:D09 | CH03MAH   |
| 978    | 16790   | RTA00002914F.c.03.1.P.Seq | F            | M00028067A:C11 | CH08LNL   |
| 979    | 139516  | RTA00002903F.l.02.1.P.Seq | F            | M00007032C:A12 | CH02COH   |
| 980    | 4825    | RTA00002930F.b.15.1.P.Seq | F            | M00042742B:E04 | CH15CON   |
| 981    | 8830    | RTA00002930F.a.05.1.P.Seq | F            | M00042528C:H01 | CH15CON   |
| 982    | 12398   | RTA00002935F.o.19.1.P.Seq | F            | M00055391B:C07 | CH17COHLV |
| 983    | 17867   | RTA00002900F.c.14.1.P.Seq | F            | M00004850A:B02 | CH02COH   |
| 984    | 15796   | RTA00002935F.b.12.1.P.Seq | F            | M00043339C:F11 | CH17COHLV |
| 985    | 185669  | RTA00002935F.f.13.1.P.Seq | F            | M00054686A:A09 | CH17COHLV |
| 986    | 13638   | RTA00002935F.i.20.1.P.Seq | F            | M00055002B:308 | CH17COHLV |



| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 987    | 8280    | RTA00002930F.e.12.1.P.Seq | F            | M00055653C:B07 | CH15CON   |
| 988    | 12632   | RTA00002931F.c.03.1.P.Seq | F            | M00042860B:C07 | CH16COP   |
| 989    | 7620    | RTA00002935F.m.18.1.P.Seq | F            | M00055240A:A08 | CH17COHLV |
| 990    | 29322   | RTA00002935F.m.20.1.P.Seq | F            | M00055244B:F07 | CH17DOHLV |
| 991    | 43864   | RTA00002931F.b.05.1.P.Seq | F            | M00042794A:F01 | CH16COP   |
| 992    | 34478   | RTA00002929F.g.13.1.P.Seq | F            | M00040367A:C08 | CH14EDT   |
| 993    | 6861    | RTA00002933F.c.17.1.P.Seq | F            | M00043221D:C12 | CH19COP   |
| 994    | 13971   | RTA00002933F.b.01.1.P.Seq | F            | M00043099A:H04 | CH19COP   |
| 995    | 13971   | RTA00002933F.a.24.1.P.Seq | F            | M00043099A:H04 | CH19COP   |
| 996    | 13244   | RTA00002927F.e.08.1.P.Seq | F            | M00039537A:F08 | CH12EDT   |
| 997    | 7455    | RTA00002935F.d.11.1.P.Seq | F            | M00054528B:E05 | CH17COHLV |
| 998    | 18915   | RTA00002929F.b.21.1.P.Seq | F            | M00040201A:H01 | CH14EDT   |
| 999    | 4023    | RTA00002935F.h.03.1.P.Seq | F            | M00054786C:D08 | CH17COHLV |
| 1000   | 10785   | RTA00002933F.a.11.1.P.Seq | F            | M00043074C:D07 | CH19COP   |
| 1001   | 14851   | RTA00002928F.i.07.1.P.Seq | F            | M00040287C:F10 | CH13EDT   |
| 1002   | 109382  | RTA00002907F.k.13.2.P.Seq | F            | M00022224A:G07 | CH03MAH   |
| 1003   | 23878   | RTA00002933F.b.03.1.P.Seq | F            | M00043101D:G11 | CH19COP   |
| 1004   | 27516   | RTA00002927F.f.17.1.P.Seq | F            | M00039598A:E04 | CH12EDT   |
| 1005   | 9652    | RTA00002931F.c.04.1.P.Seq | F            | M00042863D:F09 | CH16COP   |
| 1006   | 24729   | RTA00002931F.a.12.1.P.Seq | F            | M00042462B:C02 | CH16COP   |
| 1007   | 186041  | RTA00002912F.b.24.1.P.Seq | F            | M00027244C:B06 | CH04MAL   |
| 1008   | 12282   | RTA00002935F.i.18.1.P.Seq | F            | M00054908C:A01 | CH17COHLV |
| 1009   | 10704   | RTA00002930F.f.12.1.P.Seq | F            | M00055757A:B01 | CH15CON   |
| 1010   | 3397    | RTA00002930F.h.08.1.P.Seq | F            | M00055975B:F09 | CH15CON   |
| 1011   | 35256   | RTA00002886F.m.16.1.P.Seq | F            | M00001374A:B02 | CH01COH   |
| 1012   | 1448    | RTA00002900F.g.05.1.P.Seq | F            | M00005002A:C03 | CH02COH   |
| 1013   | 1259    | RTA00002922F.n.08.1.P.Seq | F            | M00039131C:B09 | CH09LNL   |
| 1014   | 16903   | RTA00002935F.a.01.1.P.Seq | F            | M00042352B:A04 | CH17COHLV |
| 1015   | 7884    | RTA00002922F.a.13.1.P.Seq | F            | M00038390B:F02 | CH09LNL   |
| 1016   | 5976    | RTA00002930F.j.11.1.P.Seq | F            | M00056220D:G02 | CH15CON   |
| 1017   | 16832   | RTA00002888F.h.20.1.P.Seq | F            | M00001466B:F03 | CH01COH   |
| 1018   | 10490   | RTA00002886F.k.07.1.P.Seq | F            | M00001364C:H10 | CH01COH   |
| 1019   | 6317    | RTA00002928F.g.10.1.P.Seq | F            | M00040252C:G05 | CH13EDT   |
| 1020   | 41215   | RTA00002897F.l.11.1.P.Seq | F            | M00004282A:D01 | CH01COH   |
| 1021   | 6844    | RTA00002889F.l.21.1.P.Seq | F            | M00001562B:B02 | CH01COH   |
| 1022   | 10456   | RTA00002897F.f.20.1.P.Seq | F            | M00004242D:H01 | CH01COH   |
| 1023   | 2720    | RTA00002891F.i.07.1.P.Seq | F            | M00003753A:C11 | CH01COH   |
| 1024   | 12473   | RTA00002888F.o.07.1.P.Seq | F            | M00001497C:F10 | CH01COH   |
| 1025   | 15840   | RTA00002919F.n.17.1.P.Seq | F            | M00033230C:G10 | CH08LNL   |
| 1026   | 6554    | RTA00002895F.b.16.1.P.Seq | F            | M00004062D:A02 | CH01COH   |
| 1027   | 7330    | RTA00002918F.n.17.1.P.Seq | F            | M00032987B:F01 | CH08LNL   |
| 1028   | 2206    | RTA00002919F.f.12.1.P.Seq | F            | M00033071C:G05 | CH08LNL   |
| 1029   | 42705   | RTA00002935F.f.02.1.P.Seq | F            | M00054643D:F07 | CH17COHLV |
| 1030   | 33865   | RTA00002930F.b.07.1.P.Seq | F            | M00042722C:C09 | CH15CON   |
| 1031   | 5196    | RTA00002925F.o.19.1.P.Seq | F            | M00040034B:G02 | CH09LNL   |
| 1032   | 8087    | RTA00002935F.o.10.1.P.Seq | F            | M00055375C:F12 | CH17COHLV |
| 1033   | 20072   | RTA00002935F.g.06.1.P.Seq | F            | M00054744C:F12 | CH17COHLV |
| 1034   | 12797   | RTA00002935F.g.24.1.P.Seq | F            | M00054781D:A11 | CH17COHLV |
| 1035   | 3207    | RTA00002930F.c.04.1.P.Seq | F            | M00054793B:A06 | CH15CON   |
| 1036   | 19600   | RTA00002929F.f.24.1.P.Seq | F            | M00040351D:G07 | CH14EDT   |
| 1037   | 6278    | RTA00002935F.i.18.1.P.Seq | F            | M00055001C:G10 | CH17COHLV |
| 1038   | 19363   | RTA00002927F.e.12.1.P.Seq | F            | M00039564D:D04 | CH12EDT   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 1039   | 15447   | RTA00002929F.a.12.1.P.Seq | F            | M00040366B:H10 | CH14EDT   |
| 1040   | 9676    | RTA00002932F.a.14.1.P.Seq | F            | M00042960B:C06 | CH18CON   |
| 1041   | 12560   | RTA00002929F.h.06.1.P.Seq | F            | M00040381A:B06 | CH14EDT   |
| 1042   | 12727   | RTA00002933F.c.15.1.P.Seq | F            | M00043219C:C02 | CH19COP   |
| 1043   | 27475   | RTA00002914F.c.16.1.P.Seq | F            | M00028070D:C03 | CH08LNL   |
| 1044   | 30646   | RTA00002908F.f.11.1.P.Seq | F            | M00022416D:D01 | CH03MAH   |
| 1045   | 45585   | RTA00002925F.h.20.1.P.Seq | F            | M00039894C:D09 | CH09LNL   |
| 1046   | 25025   | RTA00002925F.e.18.1.P.Seq | F            | M00039860B:E01 | CH09LNL   |
| 1047   | 15715   | RTA00002919F.p.05.1.P.Seq | F            | M00033274D:F03 | CH08LNL   |
| 1048   | 38185   | RTA00002926F.c.07.2.P.Seq | F            | M00040078A:C07 | CH09LNL   |
| 1049   | 8384    | RTA00002903F.o.13.1.P.Seq | F            | M00007112D:D03 | CH02COH   |
| 1050   | 8843    | RTA00002917F.h.17.1.P.Seq | F            | M00032733B:F12 | CH08LNL   |
| 1051   | 21401   | RTA00002930F.e.20.2.P.Seq | F            | M00055676A:G02 | CH15CON   |
| 1052   | 14434   | RTA00002903F.l.01.1.P.Seq | F            | M00007032A:B05 | CH02COH   |
| 1053   | 40045   | RTA00002891F.k.14.1.P.Seq | F            | M00003764A:H09 | CH01COH   |
| 1054   | 21853   | RTA00002896F.g.12.1.P.Seq | F            | M00004159C:D10 | CH01COH   |
| 1055   | 23439   | RTA00002935F.p.01.1.P.Seq | F            | M00055402A:H01 | CH17COHLV |
| 1056   | 13060   | RTA00002934F.a.19.1.P.Seq | F            | M00043529A:B08 | CH20COHLV |
| 1057   | 23439   | RTA00002935F.o.24.1.P.Seq | F            | M00055402A:H01 | CH17COHLV |
| 1058   | 20547   | RTA00002931F.b.12.1.P.Seq | F            | M00042822A:H04 | CH16COP   |
| 1059   | 4319    | RTA00002930F.a.03.1.P.Seq | F            | M00042525B:H01 | CH15CON   |
| 1060   | 21430   | RTA00002901F.c.18.1.P.Seq | F            | M00005452B:G03 | CH02COH   |
| 1061   | 7668    | RTA00002935F.g.23.1.P.Seq | F            | M00054781B:H04 | CH17COHLV |
| 1062   | 16239   | RTA00002935F.k.19.1.P.Seq | F            | M00055081A:A05 | CH17COHLV |
| 1063   | 5631    | RTA00002929F.e.16.1.P.Seq | F            | M00040326B:G09 | CH14EDT   |
| 1064   | 18362   | RTA00002928F.a.04.1.P.Seq | F            | M00039739B:H12 | CH13EDT   |
| 1065   | 8034    | RTA00002932F.a.22.1.P.Seq | F            | M00042982D:A10 | CH18CON   |
| 1066   | 12497   | RTA00002928F.a.19.1.P.Seq | F            | M00040132A:H09 | CH13EDT   |
| 1067   | 21001   | RTA00002932F.b.07.1.P.Seq | F            | M00042996B:H08 | CH18CON   |
| 1068   | 471     | RTA00002927F.a.11.1.P.Seq | F            | M00039184D:H09 | CH12EDT   |
| 1069   | 10003   | RTA00002897F.b.13.1.P.Seq | F            | M00004215B:C05 | CH01COH   |
| 1070   | 16074   | RTA00002935F.f.18.1.P.Seq | F            | M00054708C:B06 | CH17COHLV |
| 1071   | 13698   | RTA00002902F.l.01.1.P.Seq | F            | M00006743A:H11 | CH02COH   |
| 1072   | 24819   | RTA00002922F.i.03.1.P.Seq | F            | M00039078B:B03 | CH09LNL   |
| 1073   | 21511   | RTA00002892F.i.01.1.P.Seq | F            | M00003821C:E12 | CH01COH   |
| 1074   | 12402   | RTA00002929F.d.15.1.P.Seq | F            | M00040314B:D07 | CH14EDT   |
| 1075   | 142755  | RTA00002903F.p.06.1.P.Seq | F            | M00007126A:A02 | CH02COH   |
| 1076   | 3010    | RTA00002935F.p.12.1.P.Seq | F            | M00055423C:G12 | CH17COHLV |
| 1077   | 17173   | RTA00002935F.k.09.1.P.Seq | F            | M00055043B:H08 | CH17COHLV |
| 1078   | 2969    | RTA00002933F.a.12.1.P.Seq | F            | M00043076D:A02 | CH19COP   |
| 1079   | 19600   | RTA00002929F.g.01.1.P.Seq | F            | M00040351D:G07 | CH14EDT   |
| 1080   | 8542    | RTA00002927F.h.24.1.P.Seq | F            | M00039647A:A02 | CH12EDT   |
| 1081   | 24795   | RTA00002927F.f.10.1.P.Seq | F            | M00039594C:B06 | CH12EDT   |
| 1082   | 19695   | RTA00002927F.i.03.1.P.Seq | F            | M00039647B:A02 | CH12EDT   |
| 1083   | 8542    | RTA00002927F.i.01.1.P.Seq | F            | M00039647A:A02 | CH12EDT   |
| 1084   | 21409   | RTA00002902F.e.09.1.P.Seq | F            | M00006601D:G05 | CH02COH   |
| 1085   | 186318  | RTA00002912F.h.07.1.P.Seq | F            | M00027363D:G04 | CH04MAL   |
| 1086   | 7379    | RTA00002901F.e.10.1.P.Seq | F            | M00005468D:C01 | CH02COH   |
| 1087   | 91285   | RTA00002909F.l.12.1.P.Seq | F            | M00022682D:A10 | CH03MAH   |
| 1088   | 3285    | RTA00002903F.m.18.1.P.Seq | F            | M00007064D:D12 | CH02COH   |
| 1089   | 6284    | RTA00002896F.g.22.1.P.Seq | F            | M00004160D:G05 | CH01COH   |
| 1090   | 15676   | RTA00002935F.h.21.1.P.Seq | F            | M00054843A:C01 | CH17COHLV |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 1091   | 34112   | RTA00002894F.o.19.1.P.Seq | F           | M00004053D:F09 | CH01COH   |
| 1092   | 16407   | RTA00002892F.o.24.2.P.Seq | F           | M00003851B:A01 | CH01COH   |
| 1093   | 919     | RTA00002890F.f.18.1.P.Seq | F           | M00001609D:C11 | CH01COH   |
| 1094   | 59069   | RTA00002896F.k.07.1.P.Seq | F           | M00004175D:E06 | CH01COH   |
| 1095   | 31167   | RTA00002900F.f.05.1.P.Seq | F           | M00004876B:A06 | CH02COH   |
| 1096   | 23873   | RTA00002930F.i.03.1.P.Seq | F           | M00056035D:A08 | CH15CON   |
| 1097   | 15679   | RTA00002900F.h.12.1.P.Seq | F           | M00005016C:E04 | CH02COH   |
| 1098   | 31852   | RTA00002911F.o.22.1.P.Seq | F           | M00027170D:C07 | CH04MAL   |
| 1099   | 39030   | RTA00002891F.f.07.1.P.Seq | F           | M00001692C:C04 | CH01COH   |
| 1100   | 16407   | RTA00002892F.p.01.2.P.Seq | F           | M00003851B:A01 | CH01COH   |
| 1101   | 4118    | RTA00002935F.a.17.1.P.Seq | F           | M00043306D:C01 | CH17COHLV |
| 1102   | 11054   | RTA00002915F.n.07.2.P.Seq | F           | M00032504B:B10 | CH08LNH   |
| 1103   | 186640  | RTA00002918F.d.03.1.P.Seq | F           | M00032848D:B10 | CH08LNH   |
| 1104   | 9301    | RTA00002887F.f.22.1.P.Seq | F           | M00001396D:H02 | CH01COH   |
| 1105   | 13438   | RTA00002901F.a.11.1.P.Seq | F           | M00005422D:H10 | CH02COH   |
| 1106   | 23691   | RTA00002901F.g.21.1.P.Seq | F           | M00005506C:E09 | CH02COH   |
| 1107   | 32788   | RTA00002909F.m.04.1.P.Seq | F           | M00022694A:F05 | CH03MAH   |
| 1108   | 34364   | RTA00002915F.o.09.2.P.Seq | F           | M00032515A:B12 | CH08LNH   |
| 1109   | 24840   | RTA00002908F.i.01.1.P.Seq | F           | M00022452B:E06 | CH03MAH   |
| 1110   | 3416    | RTA00002911F.i.17.1.P.Seq | F           | M00027036A:B06 | CH04MAL   |
| 1111   | 16889   | RTA00002930F.f.04.1.P.Seq | F           | M00055724B:E04 | CH15CON   |
| 1112   | 2159    | RTA00002929F.f.15.1.P.Seq | F           | M00040344C:D05 | CH14EDT   |
| 1113   | 8880    | RTA00002929F.f.02.1.P.Seq | F           | M00040338A:B10 | CH14EDT   |
| 1114   | 10722   | RTA00002921F.o.22.1.P.Seq | F           | M00038304B:E02 | CH09LNL   |
| 1115   | 15046   | RTA00002887F.c.08.1.P.Seq | F           | M00001389B:E10 | CH01COH   |
| 1116   | 13868   | RTA00002898F.g.18.1.P.Seq | F           | M00004347B:E04 | CH01COH   |
| 1117   | 4226    | RTA00002925F.f.15.1.P.Seq | F           | M00039869A:H01 | CH09LNL   |
| 1118   | 90435   | RTA00002909F.l.06.1.P.Seq | F           | M00022678B:C08 | CH03MAH   |
| 1119   | 25686   | RTA00002900F.l.02.1.P.Seq | F           | M00005359B:D09 | CH02COH   |
| 1120   | 7296    | RTA00002900F.d.08.1.P.Seq | F           | M00004856D:F09 | CH02COH   |
| 1121   | 11546   | RTA00002905F.o.16.1.P.Seq | F           | M00021678A:H03 | CH03MAH   |
| 1122   | 15748   | RTA00002901F.h.12.1.P.Seq | F           | M00005515D:F02 | CH02COH   |
| 1123   | 5591    | RTA00002903F.p.20.1.P.Seq | F           | M00007141C:B05 | CH02COH   |
| 1124   | 9433    | RTA00002935F.p.14.1.P.Seq | F           | M00055424B:H06 | CH17COHLV |
| 1125   | 9654    | RTA00002924F.i.09.1.P.Seq | F           | M00039654C:C11 | CH09LNL   |
| 1126   | 291914  | RTA00002889F.i.06.1.P.Seq | F           | M00001550A:H06 | CH01COH   |
| 1127   | 4277    | RTA00002927F.h.13.1.P.Seq | F           | M00039642A:A08 | CH12EDT   |
| 1128   | 12362   | RTA00002929F.h.24.1.P.Seq | F           | M00040391A:G05 | CH14EDT   |
| 1129   | 449     | RTA00002924F.e.22.1.P.Seq | F           | M00039471D:G10 | CH09LNL   |
| 1130   | 1820    | RTA00002922F.n.11.1.P.Seq | F           | M00039133C:F12 | CH09LNL   |
| 1131   | 12159   | RTA00002930F.b.24.1.P.Seq | F           | M00042894C:A11 | CH15CON   |
| 1132   | 25106   | RTA00002903F.d.21.1.P.Seq | F           | M00006907B:C06 | CH02COH   |
| 1133   | 2245    | RTA00002917F.b.20.1.P.Seq | F           | M00032676C:C10 | CH08LNH   |
| 1134   | 14388   | RTA00002894F.h.09.1.P.Seq | F           | M00003998B:G10 | CH01COH   |
| 1135   | 12219   | RTA00002898F.d.22.1.P.Seq | F           | M00004328A:D01 | CH01COH   |
| 1136   | 4726    | RTA00002935F.d.16.1.P.Seq | F           | M00054538D:C12 | CH17COHLV |
| 1137   | 19479   | RTA00002891F.e.15.1.P.Seq | F           | M00001688B:B11 | CH01COH   |
| 1138   | 13280   | RTA00002888F.h.08.1.P.Seq | F           | M00001465C:A02 | CH01COH   |
| 1139   | 42708   | RTA00002901F.h.07.1.P.Seq | F           | M00005511A:F05 | CH02COH   |
| 1140   | 2022    | RTA00002896F.f.09.1.P.Seq | F           | M00004155C:A10 | CH01COH   |
| 1141   | 7281    | RTA00002929F.f.22.1.P.Seq | F           | M00040351A:C08 | CH14EDT   |
| 1142   | 3241    | RTA00002919F.n.09.1.P.Seq | F           | M00033223C:G04 | CH08LNH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTA-TION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|--------------|----------------|-----------|
| 1143   | 16161   | RTA00002930F.h.02.1.P.Seq | F            | M00055925D:B07 | CH15CON   |
| 1144   | 2766    | RTA00002935F.p.21.1.P.Seq | F            | M00055477D:B01 | CH17COHLV |
| 1145   | 11175   | RTA00002886F.b.05.1.P.Seq | F            | M00001340D:F07 | CH01COH   |
| 1146   | 7223    | RTA00002923F.i.06.1.P.Seq | F            | M00039278C:D03 | CH09LNL   |
| 1147   | 6786    | RTA00002917F.o.05.1.P.Seq | F            | M00032792C:B01 | CH08LNL   |
| 1148   | 186651  | RTA00002923F.d.21.1.P.Seq | F            | M00039219B:C08 | CH09LNL   |
| 1149   | 7878    | RTA00002930F.f.05.1.P.Seq | F            | M00055724D:C07 | CH15CON   |
| 1150   | 12624   | RTA00002935F.p.11.1.P.Seq | F            | M00055423A:B08 | CH17COHLV |
| 1151   | 23018   | RTA00002886F.f.21.1.P.Seq | F            | M00001351B:E11 | CH01COH   |
| 1152   | 186756  | RTA00002911F.g.23.1.P.Seq | F            | M00026961A:B06 | CH04MAL   |
| 1153   | 4177    | RTA00002902F.h.07.1.P.Seq | F            | M00006678A:A03 | CH02COH   |
| 1154   | 10430   | RTA00002894F.g.21.1.P.Seq | F            | M00003996B:H07 | CH01COH   |
| 1155   | 31280   | RTA00002903F.k.08.1.P.Seq | F            | M00007007A:E04 | CH02COH   |
| 1156   | 19098   | RTA00002925F.e.23.1.P.Seq | F            | M00039861C:B12 | CH09LNL   |
| 1157   | 24105   | RTA00002932F.a.06.1.P.Seq | F            | M00042585D:E10 | CH18CON   |
| 1158   | 7750    | RTA00002935F.i.02.1.P.Seq | F            | M00054856C:D03 | CH17COHLV |
| 1159   | 14582   | RTA00002898F.d.07.1.P.Seq | F            | M00004324A:D05 | CH01COH   |
| 1160   | 21356   | RTA00002917F.i.19.1.P.Seq | F            | M00032753A:C07 | CH08LNL   |
| 1161   | 16210   | RTA00002930F.k.17.1.P.Seq | F            | M00056345D:A04 | CH15CON   |
| 1162   | 2012    | RTA00002911F.o.10.1.P.Seq | F            | M00027159C:F07 | CH04MAL   |
| 1163   | 5391    | RTA00002909F.p.21.1.P.Seq | F            | M00022738D:G08 | CH03MAH   |
| 1164   | 10172   | RTA00002886F.a.05.1.P.Seq | F            | M00001338C:B02 | CH01COH   |
| 1165   | 16403   | RTA00002935F.p.15.1.P.Seq | F            | M00055424D:G05 | CH17COHLV |
| 1166   | 21920   | RTA00002886F.i.05.1.P.Seq | F            | M00001361A:C12 | CH01COH   |
| 1167   | 7070    | RTA00002921F.e.06.1.P.Seq | F            | M00033374D:C07 | CH09LNL   |
| 1168   | 45734   | RTA00002901F.i.14.1.P.Seq | F            | M00005569D:G09 | CH02COH   |
| 1169   | 12362   | RTA00002929F.i.01.1.P.Seq | F            | M00040391A:G05 | CH14EDT   |
| 1170   | 9405    | RTA00002892F.k.04.1.P.Seq | F            | M00003830C:D02 | CH01COH   |
| 1171   | 6507    | RTA00002922F.o.05.1.P.Seq | F            | M00039140A:F05 | CH09LNL   |
| 1172   | 10735   | RTA00002925F.b.24.1.P.Seq | F            | M00039822A:H02 | CH09LNL   |
| 1173   | 21177   | RTA00002935F.d.18.1.P.Seq | F            | M00054542B:A10 | CH17COHLV |
| 1174   | 14950   | RTA00002894F.m.18.1.P.Seq | F            | M00004047D:F12 | CH01COH   |
| 1175   | 10762   | RTA00002917F.o.08.1.P.Seq | F            | M00032793A:G06 | CH08LNL   |
| 1176   | 23170   | RTA00002887F.f.15.1.P.Seq | F            | M00001396B:B01 | CH01COH   |
| 1177   | 8487    | RTA00002887F.f.16.1.P.Seq | F            | M00001396B:B12 | CH01COH   |
| 1178   | 185798  | RTA00002911F.k.06.1.P.Seq | F            | M00027050A:B02 | CH04MAL   |
| 1179   | 8976    | RTA00002896F.h.03.1.P.Seq | F            | M00004161B:G07 | CH01COH   |
| 1180   | 12159   | RTA00002930F.c.01.1.P.Seq | F            | M00042894C:A11 | CH15CON   |
| 1181   | 7788    | RTA00002932F.b.13.1.P.Seq | F            | M00043017C:D08 | CH18CON   |
| 1182   | 43336   | RTA00002917F.d.09.1.P.Seq | F            | M00032688C:A03 | CH08LNL   |
| 1183   | 10313   | RTA00002902F.k.19.1.P.Seq | F            | M00006740B:A09 | CH02COH   |
| 1184   | 4588    | RTA00002891F.o.11.1.P.Seq | F            | M00003782A:B02 | CH01COH   |
| 1185   | 18090   | RTA00002925F.l.17.1.P.Seq | F            | M00039981D:B01 | CH09LNL   |
| 1186   | 185994  | RTA00002911F.p.07.1.P.Seq | F            | M00027177B:D04 | CH04MAL   |
| 1187   | 166276  | RTA00002908F.h.03.1.P.Seq | F            | M00022438C:H09 | CH03MAH   |
| 1188   | 15984   | RTA00002932F.a.10.1.P.Seq | F            | M00042621C:C04 | CH18CON   |
| 1189   | 13242   | RTA00002889F.i.11.1.P.Seq | F            | M00001550D:B11 | CH01COH   |
| 1190   | 6840    | RTA00002935F.i.06.1.P.Seq | F            | M00054866B:C08 | CH17COHLV |
| 1191   | 17265   | RTA00002935F.l.04.1.P.Seq | F            | M00055108B:A02 | CH17COHLV |
| 1192   | 12542   | RTA00002933F.b.17.1.P.Seq | F            | M00043152C:B10 | CH19COP   |
| 1193   | 1568    | RTA00002928F.d.10.1.P.Seq | F            | M00040174D:G06 | CH13EDT   |
| 1194   | 8721    | RTA00002901F.k.23.1.P.Seq | F            | M00005606D:B12 | CH02COH   |

| SEQ ID | CLUSTER | SEQ NAME                   | ORIENTA-TION | CLONE ID        | LIBRARY   |
|--------|---------|----------------------------|--------------|-----------------|-----------|
| 1195   | 13519   | RTA000002898F.i.19.1.P.Seq | F            | M000004368A:B11 | CH01COH   |
| 1196   | 4471    | RTA000002890F.d.14.1.P.Seq | F            | M00001600B:G01  | CH01COH   |
| 1197   | 11357   | RTA000002931F.c.06.1.P.Seq | F            | M00042878D:F05  | CH16COP   |
| 1198   | 11804   | RTA000002935F.m.17.1.P.Seq | F            | M00055239D:F11  | CH17COHLV |
| 1199   | 6999    | RTA000002896F.c.21.1.P.Seq | F            | M00004146A:C11  | CH01COH   |
| 1200   | 4408    | RTA000002897F.a.02.1.P.Seq | F            | M00004207C:A04  | CH01COH   |
| 1201   | 4618    | RTA000002935F.i.21.1.P.Seq | F            | M00055004C:H05  | CH17COHLV |
| 1202   | 185841  | RTA000002912F.o.08.1.P.Seq | F            | M00027600B:C07  | CH04MAL   |
| 1203   | 1278    | RTA000002912F.i.15.1.P.Seq | F            | M00027459C:B10  | CH04MAL   |
| 1204   | 19677   | RTA000002929F.h.16.1.P.Seq | F            | M00040384B:E04  | CH14EDT   |
| 1205   | 17539   | RTA000002909F.c.06.1.P.Seq | F            | M00022563B:C08  | CH03MAH   |
| 1206   | 11390   | RTA000002892F.k.03.1.P.Seq | F            | M00003830B:C06  | CH01COH   |
| 1207   | 10735   | RTA000002925F.c.01.1.P.Seq | F            | M00039822A:H02  | CH09LNL   |
| 1208   | 3239    | RTA000002887F.i.07.1.P.Seq | F            | M00001406D:F06  | CH01COH   |
| 1209   | 181718  | RTA000002905F.n.20.1.P.Seq | F            | M00021668D:A03  | CH03MAH   |
| 1210   | 6957    | RTA000002917F.n.18.1.P.Seq | F            | M00032787D:C05  | CH08LNL   |
| 1211   | 23673   | RTA000002930F.a.14.1.P.Seq | F            | M00042557D:B06  | CH15CON   |
| 1212   | 11405   | RTA000002918F.b.13.1.P.Seq | F            | M00032831A:E09  | CH08LNL   |
| 1213   | 10256   | RTA000002888F.d.16.1.P.Seq | F            | M00001449B:H10  | CH01COH   |
| 1214   | 25563   | RTA000002891F.b.23.1.P.Seq | F            | M00001675B:D06  | CH01COH   |
| 1215   | 2669    | RTA000002886F.l.03.1.P.Seq | F            | M00001368A:B07  | CH01COH   |
| 1216   | 185877  | RTA000002911F.b.11.2.P.Seq | F            | M00023389A:G04  | CH04MAL   |
| 1217   | 1314    | RTA000002930F.c.08.1.P.Seq | F            | M00054911D:E06  | CH15CON   |
| 1218   | 25843   | RTA000002908F.o.05.1.P.Seq | F            | M00022509A:H02  | CH03MAH   |
| 1219   | 1794    | RTA000002924F.g.06.1.P.Seq | F            | M00039560C:G06  | CH09LNL   |
| 1220   | 22038   | RTA000002896F.a.20.1.P.Seq | F            | M00004136C:B12  | CH01COH   |
| 1221   | 6011    | RTA000002924F.f.12.1.P.Seq | F            | M00039478C:B02  | CH09LNL   |
| 1222   | 41087   | RTA000002901F.o.06.1.P.Seq | F            | M00005675D:D09  | CH02COH   |
| 1223   | 18534   | RTA000002908F.o.16.1.P.Seq | F            | M00022512B:A09  | CH03MAH   |
| 1224   | 1444    | RTA000002922F.a.12.1.P.Seq | F            | M00038389D:D10  | CH09LNL   |
| 1225   | 1078    | RTA000002893F.p.24.1.P.Seq | F            | M00003972C:F07  | CH01COH   |
| 1226   | 8632    | RTA000002896F.h.09.1.P.Seq | F            | M00004163B:C03  | CH01COH   |
| 1227   | 105042  | RTA000002917F.o.11.1.P.Seq | F            | M00032795C:A03  | CH08LNL   |
| 1228   | 6878    | RTA000002912F.l.11.1.P.Seq | F            | M00027513D:F06  | CH04MAL   |
| 1229   | 23639   | RTA000002912F.i.05.1.P.Seq | F            | M00027381B:B04  | CH04MAL   |
| 1230   | 19635   | RTA000002896F.c.15.1.P.Seq | F            | M00004144D:B02  | CH01COH   |
| 1231   | 7217    | RTA000002926F.d.18.2.P.Seq | F            | M00040094B:C08  | CH09LNL   |
| 1232   | 4930    | RTA000002930F.g.04.1.P.Seq | F            | M00055810C:D03  | CH15CON   |
| 1233   | 16945   | RTA000002921F.o.01.1.P.Seq | F            | M00038290A:D12  | CH09LNL   |
| 1234   | 24790   | RTA000002890F.e.21.1.P.Seq | F            | M00001606D:D06  | CH01COH   |
| 1235   | 22721   | RTA000002932F.a.07.1.P.Seq | F            | M00042586A:B01  | CH18CON   |
| 1236   | 14861   | RTA000002901F.g.18.1.P.Seq | F            | M00005505B:E01  | CH02COH   |
| 1237   | 2452    | RTA000002921F.b.02.1.P.Seq | F            | M00033302B:F10  | CH09LNL   |
| 1238   | 19269   | RTA000002887F.p.11.1.P.Seq | F            | M00001430B:C01  | CH01COH   |
| 1239   | 16029   | RTA000002930F.i.17.1.P.Seq | F            | M00056244A:B06  | CH15CON   |
| 1240   | 3038    | RTA000002922F.m.04.1.P.Seq | F            | M00039121D:E07  | CH09LNL   |
| 1241   | 2933    | RTA000002922F.g.20.1.P.Seq | F            | M00039056B:G01  | CH09LNL   |
| 1242   | 15956   | RTA000002891F.j.23.1.P.Seq | F            | M00003761B:B02  | CH01COH   |
| 1243   | 15524   | RTA000002930F.g.09.1.P.Seq | F            | M00055818B:D01  | CH15CON   |
| 1244   | 21550   | RTA000002935F.f.21.1.P.Seq | F            | M00054725C:D09  | CH17COHLV |
| 1245   | 17567   | RTA000002918F.p.11.1.P.Seq | F            | M00033006A:F10  | CH08LNL   |
| 1246   | 20293   | RTA000002888F.i.20.1.P.Seq | F            | M00001477D:G09  | CH01COH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 1247   | 9520    | RTA00002927F.h.15.1.P.Seq | F           | M00039642C:F08 | CH12EDT   |
| 1248   | 2700    | RTA00002889F.e.21.1.P.Seq | F           | M00001539C:F12 | CH01COH   |
| 1249   | 25891   | RTA00002909F.p.23.1.P.Seq | F           | M00022740C:H11 | CH03MAH   |
| 1250   | 4298    | RTA00002908F.c.22.1.P.Seq | F           | M00022383C:A12 | CH03MAH   |
| 1251   | 20412   | RTA00002909F.e.02.1.P.Seq | F           | M00022590B:E05 | CH03MAH   |
| 1252   | 29413   | RTA00002906F.m.23.1.P.Seq | F           | M00022069D:C12 | CH03MAH   |
| 1253   | 12315   | RTA00002907F.l.24.2.P.Seq | F           | M00022240D:B11 | CH03MAH   |
| 1254   | 4930    | RTA00002930F.g.04.2.P.Seq | F           | M00055810C:D03 | CH15CON   |
| 1255   | 12018   | RTA00002924F.c.16.1.P.Seq | F           | M00039438B:D08 | CH09LNL   |
| 1256   | 10501   | RTA00002930F.h.23.1.P.Seq | F           | M00056024B:F09 | CH15CON   |
| 1257   | 11314   | RTA00002935F.k.03.1.P.Seq | F           | M00055023A:E11 | CH17COHLV |
| 1258   | 6426    | RTA00002927F.b.15.1.P.Seq | F           | M00039377B:H09 | CH12EDT   |
| 1259   | 2205    | RTA00002904F.c.03.1.P.Seq | F           | M00007195C:E11 | CH02COH   |
| 1260   | 6991    | RTA00002922F.i.12.1.P.Seq | F           | M00039081B:C04 | CH09LNL   |
| 1261   | 11928   | RTA00002906F.h.05.1.P.Seq | F           | M00021971C:B11 | CH03MAH   |
| 1262   | 28226   | RTA00002907F.n.20.1.P.Seq | F           | M00022262B:B06 | CH03MAH   |
| 1263   | 16059   | RTA00002935F.i.13.1.P.Seq | F           | M00054978C:F01 | CH17COHLV |
| 1264   | 2252    | RTA00002886F.k.24.1.P.Seq | F           | M00001368A:A08 | CH01COH   |
| 1265   | 4059    | RTA00002935F.f.19.1.P.Seq | F           | M00054714B:G10 | CH17COHLV |
| 1266   | 21795   | RTA00002901F.b.16.1.P.Seq | F           | M00005442A:B10 | CH02COH   |
| 1267   | 15049   | RTA00002935F.i.10.1.P.Seq | F           | M00054973B:E12 | CH17COHLV |
| 1268   | 5565    | RTA00002930F.c.02.1.P.Seq | F           | M00042908A:F09 | CH15CON   |
| 1269   | 20493   | RTA00002933F.a.14.1.P.Seq | F           | M00043077C:D12 | CH19COP   |
| 1270   | 20257   | RTA00002934F.a.14.1.P.Seq | F           | M00043495C:H05 | CH20COHLV |
| 1271   | 16392   | RTA00002899F.a.05.1.P.Seq | F           | M00004418B:A11 | CH01COH   |
| 1272   | 15797   | RTA00002930F.c.23.1.P.Seq | F           | M00055456C:H06 | CH15CON   |
| 1273   | 1811    | RTA00002891F.d.21.1.P.Seq | F           | M00001684D:E04 | CH01COH   |
| 1274   | 17503   | RTA00002935F.f.08.1.P.Seq | F           | M00054675D:G03 | CH17COHLV |
| 1275   | 14639   | RTA00002905F.h.12.1.P.Seq | F           | M00008073D:D01 | CH03MAH   |
| 1276   | 9146    | RTA00002934F.a.10.1.P.Seq | F           | M00043465C:H11 | CH20COHLV |
| 1277   | 10689   | RTA00002930F.h.19.1.P.Seq | F           | M00056004B:C05 | CH15CON   |
| 1278   | 11596   | RTA00002890F.e.23.1.P.Seq | F           | M00001607A:E04 | CH01COH   |
| 1279   | 23731   | RTA00002930F.g.18.1.P.Seq | F           | M00055873D:C02 | CH15CON   |
| 1280   | 25429   | RTA00002930F.h.10.1.P.Seq | F           | M00055980C:B04 | CH15CON   |
| 1281   | 1610    | RTA00002931F.b.24.1.P.Seq | F           | M00042858C:G11 | CH16COP   |
| 1282   | 1176    | RTA00002935F.a.10.1.P.Seq | F           | M00042457C:B06 | CH17COHLV |
| 1283   | 23578   | RTA00002930F.a.22.1.P.Seq | F           | M00042579A:D09 | CH15CON   |
| 1284   | 17238   | RTA00002932F.a.18.1.P.Seq | F           | M00042970C:B01 | CH18CON   |
| 1285   | 1610    | RTA00002931F.c.01.1.P.Seq | F           | M00042858C:G11 | CH16COP   |
| 1286   | 16366   | RTA00002932F.a.04.1.P.Seq | F           | M00042585A:H11 | CH18CON   |
| 1287   | 19709   | RTA00002932F.a.12.1.P.Seq | F           | M00042951D:G12 | CH18CON   |
| 1288   | 11027   | RTA00002932F.b.10.1.P.Seq | F           | M00043013B:E03 | CH18CON   |
| 1289   | 23451   | RTA00002933F.a.09.1.P.Seq | F           | M00042617B:E01 | CH19COP   |
| 1290   | 23731   | RTA00002930F.g.18.2.P.Seq | F           | M00055873D:C02 | CH15CON   |
| 1291   | 47898   | RTA00002911F.k.07.1.P.Seq | F           | M00027052A:E10 | CH04MAL   |
| 1292   | 32581   | RTA00002932F.b.02.1.P.Seq | F           | M00042986D:E03 | CH18CON   |
| 1293   | 42      | RTA00002934F.a.11.1.P.Seq | F           | M00043470A:C10 | CH20COHLV |
| 1294   | 1447    | RTA00002935F.d.14.1.P.Seq | F           | M00054536B:B01 | CH17COHLV |
| 1295   | 10449   | RTA00002935F.o.20.1.P.Seq | F           | M00055395D:D11 | CH17COHLV |
| 1296   | 35359   | RTA00002935F.h.11.1.P.Seq | F           | M00054817D:A11 | CH17COHLV |
| 1297   | 19657   | RTA00002935F.l.20.1.P.Seq | F           | M00055166C:D10 | CH17COHLV |
| 1298   | 12659   | RTA00002930F.i.21.1.P.Seq | F           | M00056133A:E11 | CH15CON   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 1299   | 9081    | RTA00002934F.a.22.1.P.Seq | F           | M00043640A:R01 | CH20COHLV |
| 1300   | 17084   | RTA00002935F.a.14.1.P.Seq | F           | M00042520B:H04 | CH17COHLV |
| 1301   | 11972   | RTA00002935F.b.10.1.P.Seq | F           | M00043336D:B03 | CH17COHLV |
| 1302   | 11077   | RTA00002935F.c.12.1.P.Seq | F           | M00043402B:G07 | CH17COHLV |
| 1303   | 126414  | RTA00002885F.a.01.1.P.Seq | F           | M00042350A:A05 | CH16COP   |
| 1304   | 113291  | RTA00002935F.m.15.1.P.Seq | F           | M00055232A:E08 | CH17COHLV |
| 1305   | 13224   | RTA00002935F.f.15.1.P.Seq | F           | M00054693A:E11 | CH17COHLV |
| 1306   | 14883   | RTA00002930F.k.14.1.P.Seq | F           | M00056320B:A03 | CH15CON   |
| 1307   | 13363   | RTA00002935F.a.02.1.P.Seq | F           | M00042352D:B03 | CH17COHLV |
| 1308   | 16869   | RTA00002889F.c.21.1.P.Seq | F           | M00001533C:G11 | CH01COH   |
| 1309   | 16      | RTA00002935F.a.06.1.P.Seq | F           | M00042449B:F05 | CH17COHLV |
| 1310   | 4359    | RTA00002903F.j.16.1.P.Seq | F           | M00006994C:F06 | CH02COH   |
| 1311   | 20726   | RTA00002908F.a.17.1.P.Seq | F           | M00022363C:D05 | CH03MAH   |
| 1312   | 13713   | RTA00002924F.l.09.1.P.Seq | F           | M00039686C:C01 | CH09LNL   |
| 1313   | 29271   | RTA00002935F.d.20.1.P.Seq | F           | M00054548C:H06 | CH17COHLV |
| 1314   | 6237    | RTA00002935F.f.14.1.P.Seq | F           | M00054686A:F10 | CH17COHLV |
| 1315   | 3472    | RTA00002922F.n.12.1.P.Seq | F           | M00039134D:F08 | CH09LNL   |
| 1316   | 186798  | RTA00002911F.f.11.1.P.Seq | F           | M00026914C:H09 | CH04MAL   |
| 1317   | 13193   | RTA00002886F.l.16.1.P.Seq | F           | M00001369A:G06 | CH01COH   |
| 1318   | 3554    | RTA00002919F.i.14.1.P.Seq | F           | M00033149B:E10 | CH08LNL   |
| 1319   | 19991   | RTA00002908F.h.11.1.P.Seq | F           | M00022446C:H06 | CH03MAH   |
| 1320   | 173046  | RTA00002901F.o.19.1.P.Seq | F           | M00005703D:G10 | CH02COH   |
| 1321   | 21798   | RTA00002932F.b.12.1.P.Seq | F           | M00043016B:F09 | CH18CON   |
| 1322   | 11303   | RTA00002898F.i.11.1.P.Seq | F           | M00004359A:E01 | CH01COH   |
| 1323   | 4026    | RTA00002915F.m.02.2.P.Seq | F           | M00032494C:H08 | CH08LNL   |
| 1324   | 94859   | RTA00002909F.i.23.1.P.Seq | F           | M00022656D:D07 | CH03MAH   |
| 1325   | 12315   | RTA00002907F.m.01.1.P.Seq | F           | M00022240D:B11 | CH03MAH   |
| 1326   | 4822    | RTA00002909F.l.16.1.P.Seq | F           | M00022690A:A07 | CH03MAH   |
| 1327   | 97129   | RTA00002909F.l.13.1.P.Seq | F           | M00022684A:E06 | CH03MAH   |
| 1328   | 15996   | RTA00002897F.l.09.1.P.Seq | F           | M00004281A:C04 | CH01COH   |
| 1329   | 7209    | RTA00002918F.c.01.1.P.Seq | F           | M00032835D:G04 | CH08LNL   |
| 1330   | 111888  | RTA00002902F.h.08.1.P.Seq | F           | M00006678C:C02 | CH02COH   |
| 1331   | 15642   | RTA00002902F.g.06.1.P.Seq | F           | M00006646A:A07 | CH02COH   |
| 1332   | 20016   | RTA00002916F.f.05.1.P.Seq | F           | M00032567B:G05 | CH08LNL   |
| 1333   | 21603   | RTA00002902F.a.05.1.P.Seq | F           | M00005763D:A01 | CH02COH   |
| 1334   | 156903  | RTA00002907F.i.09.2.P.Seq | F           | M00022200B:B05 | CH03MAH   |
| 1335   | 1425    | RTA00002916F.b.19.1.P.Seq | F           | M00032541C:G03 | CH08LNL   |
| 1336   | 186061  | RTA00002911F.e.24.1.P.Seq | F           | M00026900A:H07 | CH04MAL   |
| 1337   | 20717   | RTA00002907F.o.19.1.P.Seq | F           | M00022273A:E03 | CH03MAH   |
| 1338   | 12586   | RTA00002887F.a.09.1.P.Seq | F           | M00001385A:E07 | CH01COH   |
| 1339   | 19719   | RTA00002914F.h.23.1.P.Seq | F           | M00028212D:C05 | CH08LNL   |
| 1340   | 474     | RTA00002917F.g.15.1.P.Seq | F           | M00032727A:E04 | CH08LNL   |
| 1341   | 11907   | RTA00002923F.o.07.1.P.Seq | F           | M00039381C:C07 | CH09LNL   |
| 1342   | 6806    | RTA00002928F.d.02.1.P.Seq | F           | M00040169A:G06 | CH13EDT   |
| 1343   | 13146   | RTA00002892F.f.10.2.P.Seq | F           | M00003814A:G05 | CH01COH   |
| 1344   | 16686   | RTA00002919F.f.14.1.P.Seq | F           | M00033072A:A09 | CH08LNL   |
| 1345   | 6823    | RTA00002888F.a.04.1.P.Seq | F           | M00001433B:E02 | CH01COH   |
| 1346   | 43029   | RTA00002897F.d.03.1.P.Seq | F           | M00004225D:E03 | CH01COH   |
| 1347   | 14789   | RTA00002935F.k.11.1.P.Seq | F           | M00055055C:F01 | CH17COHLV |
| 1348   | 186061  | RTA00002911F.f.01.1.P.Seq | F           | M00026900A:H07 | CH04MAL   |
| 1349   | 12823   | RTA00002921F.g.24.1.P.Seq | F           | M00033434D:F05 | CH09LNL   |
| 1350   | 25844   | RTA00002908F.k.23.1.P.Seq | F           | M00022474B:C08 | CH03MAH   |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 1351   | 47793   | RTA00002930F.a.03.2.P.Seq | F           | M00055809A:B09 | CH15CON   |
| 1352   | 7695    | RTA00002892F.n.18.2.P.Seq | F           | M00003845A:C07 | CH01COH   |
| 1353   | 16997   | RTA00002922F.k.15.1.P.Seq | F           | M00039107A:E12 | CH09LNL   |
| 1354   | 25441   | RTA00002906F.i.08.1.P.Seq | F           | M00021981A:C02 | CH03MAH   |
| 1355   | 4303    | RTA00002897F.o.20.1.P.Seq | F           | M00004295D:C07 | CH01COH   |
| 1356   | 5741    | RTA00002887F.c.19.1.P.Seq | F           | M00001390D:E02 | CH01COH   |
| 1357   | 17264   | RTA00002900F.a.18.1.P.Seq | F           | M00004831C:G11 | CH02COH   |
| 1358   | 11766   | RTA00002925F.f.20.1.P.Seq | F           | M00039871C:G05 | CH09LNL   |
| 1359   | 13618   | RTA00002893F.o.15.1.P.Seq | F           | M00003963D:F01 | CH01COH   |
| 1360   | 13903   | RTA00002923F.c.18.1.P.Seq | F           | M00039204A:E09 | CH09LNL   |
| 1361   | 10673   | RTA00002927F.h.23.1.P.Seq | F           | M00039646A:E06 | CH12EDT   |
| 1362   | 17412   | RTA00002932F.b.11.1.P.Seq | F           | M00043015D:D05 | CH18CON   |
| 1363   | 2218    | RTA00002919F.a.20.1.P.Seq | F           | M00033028C:A02 | CH08LNH   |
| 1364   | 5858    | RTA00002923F.i.01.1.P.Seq | F           | M00039275B:E02 | CH09LNL   |
| 1365   | 2510    | RTA00002898F.b.14.1.P.Seq | F           | M00004316A:B03 | CH01COH   |
| 1366   | 8050    | RTA00002900F.n.04.1.P.Seq | F           | M00005383A:C11 | CH02COH   |
| 1367   | 186538  | RTA00002929F.e.18.1.P.Seq | F           | M00040329A:H05 | CH14EDT   |
| 1368   | 25427   | RTA00002935F.n.20.1.P.Seq | F           | M00055337B:C04 | CH17COHLV |
| 1369   | 24098   | RTA00002901F.a.10.1.P.Seq | F           | M00005422D:H02 | CH02COH   |
| 1370   | 123823  | RTA00002905F.h.08.1.P.Seq | F           | M00008071D:H03 | CH03MAH   |
| 1371   | 3644    | RTA00002901F.c.03.1.P.Seq | F           | M00005445D:D04 | CH02COH   |
| 1372   | 27783   | RTA00002917F.a.17.1.P.Seq | F           | M00032666A:C02 | CH08LNH   |
| 1373   | 1682    | RTA00002910F.b.03.1.P.Seq | F           | M00022801D:D09 | CH03MAH   |
| 1374   | 3200    | RTA00002887F.e.07.1.P.Seq | F           | M00001393C:F04 | CH01COH   |
| 1375   | 8442    | RTA00002917F.h.23.1.P.Seq | F           | M00032734B:E12 | CH08LNH   |
| 1376   | 15353   | RTA00002910F.e.11.1.P.Seq | F           | M00022854C:G07 | CH03MAH   |
| 1377   | 6314    | RTA00002922F.b.06.1.P.Seq | F           | M00038618D:D08 | CH09LNL   |
| 1378   | 93549   | RTA00002909F.i.14.1.P.Seq | F           | M00022662C:H04 | CH03MAH   |
| 1379   | 15496   | RTA00002906F.p.03.1.P.Seq | F           | M00022088B:H02 | CH03MAH   |
| 1380   | 16572   | RTA00002886F.k.03.1.P.Seq | F           | M00001364A:C09 | CH01COH   |
| 1381   | 74821   | RTA00002890F.p.21.1.P.Seq | F           | M00001663A:A12 | CH01COH   |
| 1382   | 11315   | RTA00002889F.d.12.1.P.Seq | F           | M00001535B:B10 | CH01COH   |
| 1383   | 10859   | RTA00002894F.c.18.1.P.Seq | F           | M00003980D:C06 | CH01COH   |
| 1384   | 15391   | RTA00002914F.f.04.1.P.Seq | F           | M00028193B:E07 | CH08LNH   |
| 1385   | 23172   | RTA00002896F.b.18.1.P.Seq | F           | M00004141B:F08 | CH01COH   |
| 1386   | 22510   | RTA00002886F.l.05.1.P.Seq | F           | M00001368A:C02 | CH01COH   |
| 1387   | 17156   | RTA00002934F.a.08.1.P.Seq | F           | M00043455B:C08 | CH20COHLV |
| 1388   | 4593    | RTA00002896F.o.18.1.P.Seq | F           | M00004200C:A04 | CH01COH   |
| 1389   | 2178    | RTA00002901F.m.08.1.P.Seq | F           | M00005626D:G11 | CH02COH   |
| 1390   | 1015    | RTA00002933F.c.11.1.P.Seq | F           | M00043213A:D05 | CH19COP   |
| 1391   | 26792   | RTA00002907F.a.18.1.P.Seq | F           | M00022103C:D05 | CH03MAH   |
| 1392   | 27830   | RTA00002921F.c.07.1.P.Seq | F           | M00033344A:B06 | CH09LNL   |
| 1393   | 14648   | RTA00002898F.i.11.1.P.Seq | F           | M00004365C:G11 | CH01COH   |
| 1394   | 12585   | RTA00002897F.i.20.1.P.Seq | F           | M00004269A:F11 | CH01COH   |
| 1395   | 15825   | RTA00002916F.d.12.1.P.Seq | F           | M00032553A:A07 | CH08LNH   |
| 1396   | 7043    | RTA00002900F.h.07.1.P.Seq | F           | M00005014B:F02 | CH02COH   |
| 1397   | 29354   | RTA00002905F.c.13.1.P.Seq | F           | M00007981C:F07 | CH03MAH   |
| 1398   | 29703   | RTA00002907F.d.24.1.P.Seq | F           | M00022144C:E12 | CH03MAH   |
| 1399   | 6811    | RTA00002913F.b.07.1.P.Seq | F           | M00027724D:D04 | CH04MAL   |
| 1400   | 12657   | RTA00002906F.b.20.1.P.Seq | F           | M00021866C:H08 | CH03MAH   |
| 1401   | 2033    | RTA00002922F.e.08.2.P.Seq | F           | M00039024D:E12 | CH09LNL   |
| 1402   | 24229   | RTA00002920F.b.04.1.P.Seq | F           | M00033329C:C02 | CH08LNH   |



| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY |
|--------|---------|---------------------------|-------------|----------------|---------|
| 1403   | 20664   | RTA00002886F.a.07.1.P.Seq | F           | M00001338C:F05 | CH01COH |
| 1404   | 3656    | RTA00002902F.f.20.1.P.Seq | F           | M00006641B:F05 | CH02COH |
| 1405   | 10998   | RTA00002931F.c.07.1.P.Seq | F           | M00042878D:G06 | CH16COP |
| 1406   | 1150    | RTA00002922F.i.14.1.P.Seq | F           | M00039081B:G07 | CH09LNL |
| 1407   | 45221   | RTA00002900F.h.06.1.P.Seq | F           | M00005013D:H05 | CH02COH |
| 1408   | 34505   | RTA00002901F.a.16.1.P.Seq | F           | M00005423C:A10 | CH02COH |
| 1409   | 8175    | RTA00002924F.f.01.1.P.Seq | F           | M00039472B:E05 | CH09LNL |
| 1410   | 8175    | RTA00002924F.e.24.1.P.Seq | F           | M00039472B:E05 | CH09LNL |
| 1411   | 19375   | RTA00002903F.n.02.1.P.Seq | F           | M00007081B:C08 | CH02COH |
| 1412   | 10866   | RTA00002929F.c.15.1.P.Seq | F           | M00040219B:B07 | CH14EDT |
| 1413   | 24166   | RTA00002891F.k.07.1.P.Seq | F           | M00003763A:B02 | CH01COH |
| 1414   | 15333   | RTA00002888F.c.12.1.P.Seq | F           | M00001442C:G12 | CH01COH |
| 1415   | 44436   | RTA00002907F.b.17.1.P.Seq | F           | M00022117C:A02 | CH03MAH |
| 1416   | 9247    | RTA00002930F.a.16.1.P.Seq | F           | M00042560C:G06 | CH15CON |
| 1417   | 12317   | RTA00002908F.g.13.1.P.Seq | F           | M00022430C:C06 | CH03MAH |
| 1418   | 11968   | RTA00002890F.i.24.1.P.Seq | F           | M00001625D:B04 | CH01COH |
| 1419   | 14181   | RTA00002908F.n.09.2.P.Seq | F           | M00022499D:D08 | CH03MAH |
| 1420   | 15359   | RTA00002909F.l.02.1.P.Seq | F           | M00022677C:C01 | CH03MAH |
| 1421   | 46675   | RTA00002916F.h.03.1.P.Seq | F           | M00032584A:D06 | CH08LNH |
| 1422   | 24898   | RTA00002903F.k.17.1.P.Seq | F           | M00007019B:E01 | CH02COH |
| 1423   | 156424  | RTA00002905F.m.22.1.P.Seq | F           | M00021653A:B02 | CH03MAH |
| 1424   | 11996   | RTA00002901F.b.24.1.P.Seq | F           | M00005445A:E07 | CH02COH |
| 1425   | 11996   | RTA00002901F.c.01.1.P.Seq | F           | M00005445A:E07 | CH02COH |
| 1426   | 4784    | RTA00002894F.e.20.1.P.Seq | F           | M00003988D:B01 | CH01COH |
| 1427   | 9120    | RTA00002914F.h.10.1.P.Seq | F           | M00028210B:H03 | CH08LNH |
| 1428   | 11295   | RTA00002890F.i.15.1.P.Seq | F           | M00001632C:A10 | CH01COH |
| 1429   | 3991    | RTA00002896F.h.05.1.P.Seq | F           | M00004162D:F02 | CH01COH |
| 1430   | 20358   | RTA00002908F.b.06.1.P.Seq | F           | M00022367D:G11 | CH03MAH |
| 1431   | 12823   | RTA00002921F.h.01.1.P.Seq | F           | M00033434D:F05 | CH09LNL |
| 1432   | 147419  | RTA00002906F.g.05.1.P.Seq | F           | M00021952B:G06 | CH03MAH |
| 1433   | 12174   | RTA00002919F.f.13.1.P.Seq | F           | M00033071D:E08 | CH08LNH |
| 1434   | 35608   | RTA00002897F.o.24.1.P.Seq | F           | M00004296B:D03 | CH01COH |
| 1435   | 2325    | RTA00002894F.g.07.1.P.Seq | F           | M00003994A:B10 | CH01COH |
| 1436   | 166261  | RTA00002908F.l.05.1.P.Seq | F           | M00022475D:C07 | CH03MAH |
| 1437   | 5713    | RTA00002920F.a.09.1.P.Seq | F           | M00033324B:F04 | CH08LNH |
| 1438   | 3624    | RTA00002910F.g.06.1.P.Seq | F           | M00022901A:C05 | CH03MAH |
| 1439   | 10305   | RTA00002909F.a.07.1.P.Seq | F           | M00022530B:C04 | CH03MAH |
| 1440   | 7768    | RTA00002910F.k.22.1.P.Seq | F           | M00022992B:G12 | CH03MAH |
| 1441   | 9847    | RTA00002908F.p.07.1.P.Seq | F           | M00022516B:C05 | CH03MAH |
| 1442   | 8583    | RTA00002887F.o.06.1.P.Seq | F           | M00001426C:F06 | CH01COH |
| 1443   | 24376   | RTA00002900F.b.07.1.P.Seq | F           | M00004836B:C02 | CH02COH |
| 1444   | 8743    | RTA00002907F.n.19.1.P.Seq | F           | M00022262A:F06 | CH03MAH |
| 1445   | 22251   | RTA00002926F.c.10.2.P.Seq | F           | M00040079B:F06 | CH09LNL |
| 1446   | 12337   | RTA00002928F.d.07.1.P.Seq | F           | M00040173D:A04 | CH13EDT |
| 1447   | 13623   | RTA00002911F.d.08.2.P.Seq | F           | M00026842B:A01 | CH04MAL |
| 1448   | 5521    | RTA00002887F.i.06.1.P.Seq | F           | M00001406B:H09 | CH01COH |
| 1449   | 2193    | RTA00002933F.a.13.1.P.Seq | F           | M00043077B:F11 | CH19COP |
| 1450   | 773     | RTA00002889F.i.02.1.P.Seq | F           | M00001551D:H09 | CH01COH |
| 1451   | 142367  | RTA00002927F.h.11.1.P.Seq | F           | M00039630D:B07 | CH12EDT |
| 1452   | 19284   | RTA00002889F.e.10.1.P.Seq | F           | M00001537B:H10 | CH01COH |
| 1453   | 24011   | RTA00002924F.c.17.1.P.Seq | F           | M00039440C:G06 | CH09LNL |
| 1454   | 5930    | RTA00002911F.f.08.1.P.Seq | F           | M00026910B:G06 | CH04MAL |

| SEQ ID | CLUSTER | SEQ NAME                  | ORIENTATION | CLONE ID       | LIBRARY   |
|--------|---------|---------------------------|-------------|----------------|-----------|
| 1455   | 21581   | RTA00002902F.c.05.1.P.Seq | F           | M00005822C:A04 | CH02COH   |
| 1456   | 3662    | RTA00002925F.c.07.1.P.Seq | F           | M00039826D:E04 | CH09LNL   |
| 1457   | 4873    | RTA00002930F.b.05.1.P.Seq | F           | M00042719A:G08 | CH15CON   |
| 1458   | 11214   | RTA00002896F.h.01.1.P.Seq | F           | M00004161A:E08 | CH01COH   |
| 1459   | 22888   | RTA00002892F.l.09.1.P.Seq | F           | M00003837C:D10 | CH01COH   |
| 1460   | 15490   | RTA00002926F.k.08.1.P.Seq | F           | M00039932B:A07 | CH09LNL   |
| 1461   | 112819  | RTA00002905F.o.13.1.P.Seq | F           | M00021676C:G03 | CH03MAH   |
| 1462   | 19688   | RTA00002896F.l.02.1.P.Seq | F           | M00004179D:A12 | CH01COH   |
| 1463   | 15132   | RTA00002922F.n.20.1.P.Seq | F           | M00039138B:G05 | CH09LNL   |
| 1464   | 25022   | RTA00002914F.i.21.1.P.Seq | F           | M00028219B:H05 | CH08LNL   |
| 1465   | 16303   | RTA00002888F.b.12.1.P.Seq | F           | M00001438A:E01 | CH01COH   |
| 1466   | 16828   | RTA00002897F.b.04.1.P.Seq | F           | M00004214A:E05 | CH01COH   |
| 1467   | 14295   | RTA00002921F.a.18.1.P.Seq | F           | M00033296C:C11 | CH09LNL   |
| 1468   | 1979    | RTA00002930F.f.06.1.P.Seq | F           | M00055725D:D09 | CH15CON   |
| 1469   | 36248   | RTA00002888F.g.05.1.P.Seq | F           | M00001460C:E10 | CH01COH   |
| 1470   | 5676    | RTA00002926F.b.22.2.P.Seq | F           | M00040075B:A05 | CH09LNL   |
| 1471   | 1239    | RTA00002887F.o.21.1.P.Seq | F           | M00001428B:C10 | CH01COH   |
| 1472   | 7937    | RTA00002917F.g.22.1.P.Seq | F           | M00032728D:F01 | CH08LNL   |
| 1473   | 4483    | RTA00002911F.d.22.2.P.Seq | F           | M00026856B:G03 | CH04MAL   |
| 1474   | 7796    | RTA00002925F.c.05.1.P.Seq | F           | M00039826B:F09 | CH09LNL   |
| 1475   | 17330   | RTA00002915F.a.03.1.P.Seq | F           | M00028616C:D09 | CH08LNL   |
| 1476   | 25620   | RTA00002902F.f.09.1.P.Seq | F           | M00006631C:A04 | CH02COH   |
| 1477   | 20601   | RTA00002923F.l.20.1.P.Seq | F           | M00039326A:G07 | CH09LNL   |
| 1478   | 6205    | RTA00002923F.g.21.1.P.Seq | F           | M00039258C:C01 | CH09LNL   |
| 1479   | 726     | RTA00002913F.b.16.1.P.Seq | F           | M00027734D:C03 | CH04MAL   |
| 1480   | 104999  | RTA00002908F.g.17.1.P.Seq | F           | M00022435B:G12 | CH03MAH   |
| 1481   | 30321   | RTA00002919F.o.17.1.P.Seq | F           | M00033264B:E06 | CH08LNL   |
| 1482   | 5878    | RTA00002913F.a.16.1.P.Seq | F           | M00027688C:C01 | CH04MAL   |
| 1483   | 5944    | RTA00002905F.m.07.1.P.Seq | F           | M00021649B:A02 | CH03MAH   |
| 1484   | 5796    | RTA00002908F.i.21.1.P.Seq | F           | M00022457A:G05 | CH03MAH   |
| 1485   | 3804    | RTA00002935F.m.24.1.P.Seq | F           | M00055254A:H03 | CH17COHLV |
| 1486   | 2728    | RTA00002918F.a.22.1.P.Seq | F           | M00032828A:A06 | CH08LNL   |
| 1487   | 3804    | RTA00002935F.n.01.1.P.Seq | F           | M00055254A:H03 | CH17COHLV |
| 1488   | 3932    | RTA00002915F.o.19.2.P.Seq | F           | M00032517C:E10 | CH08LNL   |
| 1489   | 16691   | RTA00002891F.o.03.1.P.Seq | F           | M00003780A:G01 | CH01COH   |
| 1490   | 15430   | RTA00002900F.g.10.1.P.Seq | F           | M00005003D:C02 | CH02COH   |
| 1491   | 5637    | RTA00002925F.b.18.1.P.Seq | F           | M00039820B:F06 | CH09LNL   |
| 1492   | 16633   | RTA00002897F.g.15.1.P.Seq | F           | M00004246B:H07 | CH01COH   |
| 1493   | 21826   | RTA00002898F.g.06.1.P.Seq | F           | M00004344A:G11 | CH01COH   |
| 1494   | 22193   | RTA00002919F.i.09.1.P.Seq | F           | M00033146D:A03 | CH08LNL   |
| 1495   | 10720   | RTA00002898F.c.14.1.P.Seq | F           | M00004320C:E07 | CH01COH   |
| 1496   | 22491   | RTA00002925F.m.06.1.P.Seq | F           | M00040003A:G10 | CH09LNL   |
| 1497   | 10423   | RTA00002915F.n.13.2.P.Seq | F           | M00032507D:G08 | CH08LNL   |
| 1498   | 4953    | RTA00002916F.h.11.1.P.Seq | F           | M00032586C:B04 | CH08LNL   |
| 1499   | 185567  | RTA00002911F.p.08.1.P.Seq | F           | M00027178B:A11 | CH04MAL   |
| 1500   | 25605   | RTA00002924F.m.22.1.P.Seq | F           | M00039710B:A01 | CH09LNL   |
| 1501   | 29446   | RTA00002906F.m.24.1.P.Seq | F           | M00022070B:B04 | CH03MAH   |
| 1502   | 9668    | RTA00002908F.g.02.1.P.Seq | F           | M00022421A:F12 | CH03MAH   |
| 1503   | 29446   | RTA00002906F.n.01.1.P.Seq | F           | M00022070B:B04 | CH03MAH   |
| 1504   | 7171    | RTA00002887F.m.22.1.P.Seq | F           | M00001421B:E07 | CH01COH   |

TABLE 3

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|-------------|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 2      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 3      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 4      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 5      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 6      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 7      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 8      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 9      | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 10     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 11     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 12     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 13     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 14     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 15     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 16     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 17     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 18     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 19     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 20     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 21     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 22     | <NONE>                                | <NONE>      | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 23     | <NONE>                                | <NONE>      | <NONE>  | 548562   | GENOME POLYPROTEIN [CONTAINS: RNA REPLICASE; HELICASE; COAT PROTEIN] 2.7.7.48) - apple stem grooving virus (strain P-209) | 9.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|-------------|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 24     | <NONE>                                | <NONE>      | <NONE>  | 416959   | EXCISION REPAIR PROTEIN ERCC-6 DNA repair helicase ERCC6 - human >gi 182181 (L04791) excision repair protein [Homo sapiens]                       | 8.9     |
| 25     | <NONE>                                | <NONE>      | <NONE>  | 3327096  | (AB014541) KIAA0641 protein [Homo sapiens]  | 8.7     |
| 26     | <NONE>                                | <NONE>      | <NONE>  | 861293   | (U28741) F35D2.1 gene product [Caenorhabditis elegans]  | 7.9     |
| 27     | <NONE>                                | <NONE>      | <NONE>  | 3297821  | (AL031032) extensin-like protein  | 5.5     |
| 28     | <NONE>                                | <NONE>      | <NONE>  | 2119692  | transforming growth factor-beta type III receptor - chicken >gi 511843 (L01121) transforming growth factor-beta type III receptor [Gallus gallus] | 5.1     |
| 29     | <NONE>                                | <NONE>      | <NONE>  | 2136028  | protein kinase PRKI - human   | 5.0     |
| 30     | <NONE>                                | <NONE>      | <NONE>  | 2746912  | (AF040659) No definition line found [Caenorhabditis elegans]  | 4.6     |
| 31     | <NONE>                                | <NONE>      | <NONE>  | 2358287  | (AF010404) ALR [Homo sapiens]   | 4.5     |
| 32     | <NONE>                                | <NONE>      | <NONE>  | 3877816  | (Z96048) predicted using Genefinder; cDNA EST EMBL:D65516 comes from this gene; cDNA EST yk191a5.5 comes from this gene [Caenorhabditis elegans]  | 4.4     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|-------------|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 33     | <NONE>                                | <NONE>      | <NONE>  | 4140268  | (Y14953) SRCR domain, membrane form 2  | 4.1     |
| 34     | <NONE>                                | <NONE>      | <NONE>  | 1708663  | (U51183) transposase [Hydra vulgaris]  | 4.0     |
| 35     | <NONE>                                | <NONE>      | <NONE>  | 1184100  | (U45958) pistil extensin-like protein [Nicotiana glauca]   | 3.9     |
| 36     | <NONE>                                | <NONE>      | <NONE>  | 121073   | GLUCOCORTICOID RECEPTOR (GR)   | 3.9     |
| 37     | <NONE>                                | <NONE>      | <NONE>  | 1718298  | (U75698) ORF 45; contains an extended acidic domain; EBV BKRF4 homolog [Kaposi's sarcoma-associated herpesvirus] homolog, conserved in other gamma-herpesviruses | 2.6     |
| 38     | <NONE>                                | <NONE>      | <NONE>  | 2352538  | (AF006564) alcohol dehydrogenase [Drosophila persimilis] persimilis]   | 1.4     |
| 39     | <NONE>                                | <NONE>      | <NONE>  | 3192897  | (AF066071) SP85; PsB [Dictyostelium discoideum]  | 1.4     |
| 40     | <NONE>                                | <NONE>      | <NONE>  | 561645   | (L33421) This CDS feature is included to show the translation of the corresponding V_region. Presently translation qualifiers on V_region features are illegal   | 1.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|-------------|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 41     | <NONE>                                | <NONE>      | <NONE>  | 3878857  | (Z83120) predicated using Genefinder; cDNA EST EMBL:D35016 comes from this gene; cDNA EST EMBL:D32583 comes from this gene; cDNA EST EMBL:D35258 comes from this gene; cDNA EST EMBL:C11471 comes from this gene; cDNA EST EMBL:C ... | 1.0     |
| 42     | <NONE>                                | <NONE>      | <NONE>  | 1658571  | (U75903) UTG1A7 [Rattus norvegicus]   | 1.0     |
| 43     | <NONE>                                | <NONE>      | <NONE>  | 2338034  | (AF005370) putative immediate early protein [Alcelaphine herpesvirus 1]   | 0.86    |
| 44     | <NONE>                                | <NONE>      | <NONE>  | 3043714  | (AB011167) KIAA0595 protein [Homo sapiens]  | 0.42    |
| 45     | <NONE>                                | <NONE>      | <NONE>  | 1723710  | HYPOTHETICAL 92.7 KD PROTEIN IN ASN2-PHB1 INTERGENIC REGION<br>>gi 2131678 pir  S64439 hypothetical protein YRG130c - yeast (Saccharomyces cerevisiae)<br>>gi 1323215 gnl PID e243523 (Z72915) ORF YGR130c [Saccharomyces cerevisiae] | 0.40    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|-------------|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 46     | <NONE>                                | <NONE>      | <NONE>  | 1723710  | HYPOTHETICAL<br>92.7 KD PROTEIN<br>IN ASN2-PHB1<br>INTERGENIC<br>REGION<br>>gi 2131678 pir  S644<br>39 hypothetical<br>protein YRG130c -<br>yeast ( <i>Saccharomyces<br/>cerevisiae</i> )<br>>gi 1323215 gnl PID e<br>243523 (Z72915)<br>ORF YGR130c<br>[ <i>Saccharomyces<br/>cerevisiae</i> ] | 0.38    |
| 47     | <NONE>                                | <NONE>      | <NONE>  | 2996117  | (AF046125)<br>immediate early 2 [Rat<br>cytomegalovirus]  | 0.26    |
| 48     | <NONE>                                | <NONE>      | <NONE>  | 4151809  | (AF102855) synaptic<br>SAPAP-interacting<br>protein Synamon   | 0.024   |
| 49     | <NONE>                                | <NONE>      | <NONE>  | 2773341  | (AF040954) putative<br>protein phosphatase 1<br>nuclear targeting<br>subunit [ <i>Rattus<br/>norvegicus</i> ]   | 0.017   |
| 50     | <NONE>                                | <NONE>      | <NONE>  | 1653522  | (D90914) hypothetical<br>protein  | 3e-04   |
| 51     | <NONE>                                | <NONE>      | <NONE>  | 3219965  | HYPOTHETICAL<br>100.6 KD TRP-ASP<br>REPEATS<br>CONTAINING<br>PROTEIN C2C6.04C<br>IN CHROMOSOME I  | 3e-06   |
| 52     | <NONE>                                | <NONE>      | <NONE>  | 4185567  | (AF115480) cAMP-<br>dependent Rap1<br>guanine-nucleotide<br>exchange factor [ <i>Mus<br/>musculus</i> ]   | 7e-07   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 53     | <NONE>                                | <NONE>  | <NONE>  | 1176527  | HYPOTHETICAL 43.2 KD PROTEIN C34E10.1 IN CHROMOSOME III >gi 500724 (U10402) C34E10.1 gene product [Caenorhabditis elegans] | 3e-20   |
| 54     | X85444                                | G.pallida repetitive DNA element  | 5.0     | 2118936  | beta-globin - chimpanzee (fragment)  | 8.6     |
| 55     | X72961                                | Synechococcus sp. cpeB, cpeA genes and ORF3                                   | 5.0     | 462569   | MICROTUBULE-ASSOCIATED PROTEIN 1A microtubule-associated protein MAP1A - rat >gi 205538 norvegicus]                        | 2.2     |
| 56     | U94747                                | Human WD repeat protein HAN11 mRNA, complete cds                              | 5.0     | 3875538  | (Z67990) similar to cuticle collagen   | 1.3     |
| 57     | AF032108                              | Homo sapiens integrin alpha-7 mRNA, complete cds                              | 5.0     | 2147194  | collagen - Paralvinella grasslei   | 0.002   |
| 58     | Z50798                                | G.gallus mRNA for p52   | 5.0     | 3122885  | ASPARTYL-TRNA SYNTHETASE synthetase [Bacillus subtilis]  | 3e-11   |
| 59     | AB002384                              | Human mRNA for KIAA0386 gene, complete cds                                    | 5.0     | 2632098  | (Y15513) Prodos protein [Drosophila melanogaster]  | 9e-12   |
| 60     | X14835                                | Thermofilum pendens DNA for 16S and 23S ribosomal RNA, tRNA-Met, and tRNA Gly | 4.9     | <NONE>   | <NONE>   | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 61     | U87149                                | Hordeum vulgare nucellin gene, complete cds   | 4.9     | 128578   | NONSTRUCTURAL PROTEIN NS-S spotted wilt virus (strain CPNH1) non-structural protein [Tomato spotted wilt virus]                     | 2.8     |
| 62     | D87541                                | Mus musculus gene for integrin alpha v subunit, promoter region   | 4.9     | 136956   | HYPOTHETICAL PROTEIN UL61 cytomegalovirus (strain AD169) cytomegalovirus  | 0.038   |
| 63     | U72520                                | Mus musculus mena protein (Mena) mRNA, complete cds   | 4.9     | 3413892  | (AB007934) KIAA0465 protein [Homo sapiens]  | 6e-07   |
| 64     | S79797                                | enzymatic glycosylation-regulating gene [rats, Sprague-Dawley, streptozotocin diabetic, heart, mRNA, 5010 nt] | 4.8     | <NONE>   | <NONE>  | <NONE>  |
| 65     | AB011102                              | Homo sapiens mRNA for KIAA0530 protein, partial cds   | 4.8     | 138022   | RECEPTOR RECOGNIZING PROTEIN gp38 - phage Ox2 >gi 15126 (X05675) gene 38 (AA 1-266); pid:g15126 [Bacteriophage Ox2]                 | 3.6     |
| 66     | AF100985                              | Penaeus monodon phosphopyruvate hydratase mRNA, complete cds  | 4.8     | 500615   | (D16221) endochitinase [Oryza sativa]   | 2.8     |
| 67     | U31756                                | Bacillus subtilis gamma-aminobutyrate permease cds  | 4.8     | 3880699  | (AL021471) similar to Eukaryotic aspartyl proteases [Caenorhabditis elegans] Eukaryotic aspartyl proteases [Caenorhabditis elegans] | 2.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 68     | U25111                                | Pisum sativum chloroplast processing enzyme mRNA, nuclear gene encoding chloroplast protein, complete cds. | 4.8     | 1800145  | (U83658) FH1/FH2 protein homolog [Emericella nidulans]                  | 1.6     |
| 69     | U00454                                | Mus musculus Cdx-2 homeobox protein gene, complete cds.  | 4.7     | <NONE>   | <NONE>  | <NONE>  |
| 70     | M84166                                | Hamster c-Ha-ras protein gene, complete cds.   | 4.7     | 1710606  | RENIN-BINDING PROTEIN (RBNP) protein [Rattus norvegicus]                | 0.88    |
| 71     | AF087516                              | Mus musculus major sperm fibrous sheath protein Pro-mAKAP82 gene, alternative splice exons 1' and 1''      | 4.6     | <NONE>   | <NONE>  | <NONE>  |
| 72     | X74160                                | M.esculenta mRNA for granule-bound starch synthase   | 4.6     | <NONE>   | <NONE>  | <NONE>  |
| 73     | M97487                                | Haloferax volcanii superoxide dismutase (sod2) gene, complete cds.   | 4.6     | 2623307  | (AC002409) putative ubiquitin protease [Arabidopsis thaliana]           | 3.4     |
| 74     | M57889                                | Drosophila melanogaster suppressor of sable gene, complete cds.  | 4.5     | <NONE>   | <NONE>  | <NONE>  |
| 75     | D49708                                | Rattus norvegicus mRNA for RNA binding protein   | 4.5     | <NONE>   | <NONE>  | <NONE>  |
| 76     | D31853                                | Yeast GTS1 gene for glycyl-threonin/serine repeat protein, complete cds                                    | 4.5     | 2447195  | (U42580) NETTF (7x), DETTS (4x) [Paramecium bursaria Chlorella virus 1] | 3.3     |
| 77     | Z47036                                | Human partial cDNA sequence, clone bs613;  | 2.9     | <NONE>   | <NONE>  | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 78                                    | L19660         | Rattus norvegicus gastric inhibitory peptide receptor mRNA, complete cds                    | 2.7     | 2358279   | (AF007871) torsinA [Homo sapiens]  | 2e-07   |
| 79                                    | X82841         | A.thaliana Aco gene   | 2.6     | 483212  | immediate-early protein IE110 - human herpesvirus 1 (strain HFEM) (fragment) | 8.4     |
| 80                                    | X61931         | S.purpurascens famA and famB genes for FAS domain and acyl-CoA-dehydrogenases, respectively | 2.6     | 2290534   | (U95031) sublingual gland mucin [Homo sapiens]                               | 0.47    |
| 81                                    | U13680         | Human lactate dehydrogenase-C (LDH-C) mRNA, complete cds.                                   | 2.5     | 2887449   | (AB007874) KIAA0414 [Homo sapiens]   | 3.1     |
| 82                                    | AB007869       | Homo sapiens KIAA0409 mRNA, partial cds   | 2.4     | 3130157   | (AB008859) pheromone receptor [Fugu rubripes]                                | 3.1     |
| 83                                    | X97479         | H.sapiens mas proto-oncogene, 5' region   | 2.1     | <NONE>  | <NONE>   | <NONE>  |
| 84                                    | X98374         | R.norvegicus mRNA for KIS protein   | 1.9     | <NONE>  | <NONE>   | <NONE>  |
| 85                                    | AE000710       | Aquifex aeolicus section 42 of 109 of the complete genome                                   | 1.9     | <NONE>  | <NONE>   | <NONE>  |
| 86                                    | D30612         | Homo sapiens mRNA for repressor protein, partial cds  | 1.9     | <NONE>  | <NONE>   | <NONE>  |
| 87                                    | Y14321         | Homo sapiens PMP69 gene, exons 8,9,10 & 11  | 1.9     | <NONE>  | <NONE>   | <NONE>  |
| 88                                    | D90773         | E.coli genomic DNA, Kohara clone #262(30.3-30.5 min.)                                       | 1.9     | 1536816   | (D78305) DNA binding protein [Chlorella virus]                               | 7.9     |
| 89                                    | AE000991       | Archaeoglobus fulgidus section 116 of 172 of the complete genome                            | 1.9     | 520645  | (X79095) pyruvate, orthophosphate dikinase [Flaveria trinervia]              | 2.7     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 90                                    | U39476         | Rattus norvegicus p95 Vav (Vav) proto-oncogene mRNA, complete cds.                                 | 1.9     | 4158178   | (AL023496) hypothetical protein   | 1.6     |
| 91                                    | U28838         | Human transcription factor TFIIIB 90 kDa subunit   | 1.9     | 2495730   | HYPOTHETICAL PROLINE-RICH PROTEIN KIAA0269 >gi 1665805 gnl PID d1014089 (D87459) Similar to Volbox carteri extensin (S22697) [Homo sapiens] | 0.23    |
| 92                                    | U20106         | Rattus norvegicus synaptotagmin VII mRNA, complete cds.  | 1.9     | 478380  | UL47h protein - Marek's disease virus   | 0.23    |
| 93                                    | AF071010       | Mouse mammary tumor virus putative integrase, env polyprotein, and superantigen mRNA, complete cds | 1.9     | 2781386   | (AC004010) similar to Leucine-rich transmembrane proteins; 44% similarity to U42767 (PID:g1736918) [Homo sapiens]                           | 4e-33   |
| 94                                    | AF061881       | Mesocricetus auratus c-fos proto-oncogene protein (c-fos) gene, complete cds                       | 1.8     | <NONE>  | <NONE>  | <NONE>  |
| 95                                    | AE001397       | Plasmodium falciparum chromosome 2, section 34 of 73 of the complete sequence                      | 1.8     | <NONE>  | <NONE>  | <NONE>  |
| 96                                    | D14701         | Horseshoe crab mRNA for coagulation factor B, complete cds   | 1.8     | <NONE>  | <NONE>  | <NONE>  |
| 97                                    | M29154         | P.falciparum multidrug resistance (MRD) gene, complete cds.  | 1.8     | <NONE>  | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 98     | L16532                                | Rattus norvegicus (cline pCNPII) 2',3'-cyclic nucleotide 3'-phosphodiesterase (CNPII) mRNA, complete cds.                    | 1.8     | <NONE>   | <NONE>   | <NONE>  |
| 99     | AE001434                              | Plasmodium falciparum chromosome 2, section 71 of 73 of the complete sequence  | 1.8     | <NONE>   | <NONE>   | <NONE>  |
| 100    | Z46785                                | D.melanogaster gene for protamine (mst35Bb).   | 1.8     | <NONE>   | <NONE>   | <NONE>  |
| 101    | X69822                                | P.sylvestris mRNA for glutamine synthetase   | 1.8     | 219896   | (D90452) 1-caldesmon I [Homo sapiens]  | 9.7     |
| 102    | U49055                                | Rattus norvegicus CTD-binding SR-like protein rA8 mRNA, complete cds   | 1.8     | 2497252  | INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 (IGFBP-4) (IBP-4) (IGF-BINDING PROTEIN 4) factor-binding protein-4 - sheep (fragment) factor-binding protein-4, IGFBP-4 [sheep liver, Peptide, 237 aa] [Ovis aries] | 2.5     |
| 103    | L28101                                | Homo sapiens kallistatin (PI4) gene, exons 1-4, complete cds   | 1.8     | 4204267  | (AC005223) 55585 [Arabidopsis thaliana]  | 2.4     |
| 104    | U66987                                | Pandorina morum internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence | 1.8     | 2635909  | (Z99121) permease [Bacillus subtilis]  | 1.9     |

| Nearest Neighbor (BlastN vs. Genbank) |             |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-------------|---|---------|--|--|---------|
| SEQ ID                                | ACCES- SION | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 105                                   | X58033      | Human polymorphic MspI site DNA (D3S3 locus)              | 1.8     | 2136878  | keratin KAP5.5 - sheep (fragment) >gi 313722   | 0.65    |
| 106                                   | U15780      | Human p82 (ST5) mRNA, alternatively spliced, complete cds | 1.8     | 3638957  | (AC004877) spondin-mucin-like; similar to P98167 uncertain [Homo sapiens]  | 0.64    |
| 107                                   | AF038535    | Homo sapiens synaptotagmin VII mRNA, partial cds          | 1.8     | 457927   | (U00690) calcium channel alpha-I subunit [Drosophila melanogaster]   | 0.51    |
| 108                                   | AF052134    | Homo sapiens clone 23585 mRNA sequence                    | 1.8     | 232263   | HOMEBOX PROTEIN HOX-DI (HOX-4.9)   | 0.28    |
| 109                                   | X75208      | H.sapiens HEK2 mRNA for protein tyrosine kinase receptor. | 1.8     | 1730198  | GROWTH-ARREST-SPECIFIC PROTEIN 1 gene product [Homo sapiens]   | 0.22    |
| 110                                   | AB013896    | Xenopus laevis mRNA for SOX-D, complete cds               | 1.8     | 2494501  | TRANSCRIPTION FACTOR FKH-4 factor [Mus musculus]   | 0.17    |
| 111                                   | D16947      | Human HepG2 3' region cDNA, clone hmd6b10                 | 1.8     | 3413870  | (AB007923) KIAA0454 protein [Homo sapiens]   | 0.002   |
| 112                                   | D13547      | Mouse DNA, Tearly alpha (TEA) region                      | 1.8     | 3393018  | (AL031174) hypothetical protein  | 53-08   |
| 113                                   | M35498      | Woodchuck e-myc protein gene, exon 1.                     | 1.8     | 3183405  | HYPOTHETICAL 11.3 KD PROTEIN C2C6.07 IN CHROMOSOME I >gi 2370504 gnl PID e 1339194 pombe] >gi 3451305 gnl PID e 1316730 (AL031324) very hypothetical protein [Schizosaccharomyces pombe] | 8e-10   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION                                       | P VALUE |
| 114    | M94166                                | Hamster c-Ha-ras protein gene, complete cds.                                  | 1.8     | 3386622  | (AC004665) unknown protein [Arabidopsis thaliana] | 2e-10   |
| 115    | U33135                                | Mychodea carnosia 18S ribosomal RNA gene, complete sequence                   | 1.8     | 3334982  | (AC005306) R27216_1 [Homo sapiens]                | 3e-22   |
| 116    | U84003                                | Homo sapiens putative tumor suppressor (BIN1) gene, exons 7-12                | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 117    | AE001121                              | Borrelia burgdorferi (section 7 of 70) of the complete genome                 | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 118    | AE001114                              | Archaeoglobus fulgidus section 165 of 172 of the complete genome              | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 119    | U82064                                | Angiostrongylus cantonensis adult-specific muscle protein-1 gene, partial cds | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 120    | AF041836                              | Buchnera aphidicola plasmid pLeu-Sg, complete plasmid sequence                | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 121    | M87479                                | Lymnaea stagnalis FMRFamide gene, mature peptides.                            | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 122    | M55163                                | Xenopus laevis fibroblast growth factor receptor mRNA, complete cds.          | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 123    | S57565                                | histamine H2-receptor [rats, Genomic, 1928 nt]                                | 1.7     | <NONE>   | <NONE>  | <NONE>  |
| 124    | M27256                                | Simian immunodeficiency virus (SIV) pol region.                               | 1.7     | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION | P VALUE |
| 125    | U31516                                | Human chromosome 8 anonymous clone pBS8-165  | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 126    | X12671                                | Human gene for heterogeneous nuclear ribonucleoprotein (hnRNP) core protein A1         | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 127    | AF009054                              | Paeonia suffruticosa ssp. spontenea alcohol dehydrogenase 1B (Adh1B) gene, partial cds | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 128    | AF046917                              | Mus musculus transketolase gene, exon 6 and partial cds                                | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 129    | D89053                                | Homo sapiens mRNA for Acyl-CoA synthetase 3, complete cds                              | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 130    | U57968                                | Staphylothermus marinus surface layer-associated STABLE protease gene, complete cds.   | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 131    | L39072                                | Bovine herpesvirus 1 (clone p95) UL24 homologue gene, complete cds.                    | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 132    | X04980                                | Drosophila simulans retrotransposon 297 5'-LTR and flanks (pWK1020)                    | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 133    | AE001114                              | Archaeoglobus fulgidus section 165 of 172 of the complete genome                       | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 134    | X04434                                | Human mRNA for insulin-like growth factor I receptor                                   | 1.7     | <NONE>   | <NONE>      | <NONE>  |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |                                 |         |
|---------------------------------------|----------------|--|---------|---|---------------------------------|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION                     | P VALUE |
| 135                                   | U07890         | Mus musculus C57BL/6J epidermal surface antigen (mesa) mRNA, complete cds.   | 1.7     | <NONE>  | <NONE>                          | <NONE>  |
| 136                                   | D26163         | Human tyrosinase gene, 5'-flanking region cell-specific transcription)   | 1.7     | <NONE>  | <NONE>                          | <NONE>  |
| 137                                   | AF093818       | Panorpa nipponensis NADH dehydrogenase subunit 5 gene, mitochondrial gene encoding mitochondrial protein, partial cds  | 1.7     | <NONE>  | <NONE>                          | <NONE>  |
| 138                                   | D50560         | Xenopus laevis mRNA for cytochrome P-450, complete cds   | 1.7     | <NONE>  | <NONE>                          | <NONE>  |
| 139                                   | AF083488       | Mus musculus phospholipase D1 (PLD1) gene, exons 18 and 19, complete sequence  | 1.7     | <NONE>  | <NONE>                          | <NONE>  |
| 140                                   | AF100694       | Mus musculus Pontin52 mRNA, complete cds   | 1.7     | <NONE>  | <NONE>                          | <NONE>  |
| 141                                   | M73749         | Streptococcus salivarius thermophilus beta-D-galactose (lacZ) gene, complete cds.<br>>::gb M63636 STRLA CZZ Streptococcus thermophilus beta-D-galactosidase (lacZ) gene, complete cds. | 1.7     | <NONE>  | <NONE>                          | <NONE>  |
| 142                                   | AE001114       | Archaeoglobus fulgidus section 165 of 172 of the complete genome   | 1.7     | 2183023   | (U84971) unknown [Homo sapiens] | 9.2     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 143                                   | L01983         | Human type IV sodium channel alpha polypeptide  | 1.7     | 130504  | GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER COMPONENT PROTEINASE INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2); GENOME-LINKED PROTEIN (VPG); NUCLEAR ... virus (strain D) | 9.2     |
| 144                                   | L19731         | Plecotus rafinesquii mitochondrial cytochrome b gene, 5' end.   | 1.7     | 3327096   | (AB014541) KIAA0641 protein [Homo sapiens]   | 9.1     |
| 145                                   | AE001114       | Archaeoglobus fulgidus section 16S of 172 of the complete genome  | 1.7     | 2183023   | (U84971) unknown [Homo sapiens]  | 8.8     |
| 146                                   | L27218         | Bos taurus serum amine oxidase mRNA, complete cds. > oxidase=amiloride-binding protein homolog [cattle, liver, mRNA, 2664 nt] | 1.7     | 1174459   | SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 6 (IL-4 STAT) >gi 559855 (U16031) IL-4 Stat [Homo sapiens]  | 7.1     |
| 147                                   | Z49868         | Caenorhabditis elegans cosmid W07E11, complete sequence [Caenorhabditis elegans]  | 1.7     | 4204263   | (AC005223) 40409 [Arabidopsis thaliana]  | 6.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 148    | AL022271                              | Caenorhabditis elegans cosmid F32F2, complete sequence [Caenorhabditis elegans] | 1.7     | 2497969  | PERIPLASMIC NITRATE REDUCTASE PRECURSOR<br>>gi 1086107 pir  S50163.nitrate reductase large chain precursor, periplasmic - Thiosphaera pantotropha<br>>gi 600093 (Z36773) periplasmic nitrate reductase large subunit [Paracossus denitrificans] | 6.7     |
| 149    | U43844                                | Mus musculus cyclin D3 gene, complete cds                                       | 1.7     | 3861490  | (AF062037) capsid protein precursor [Thosea asigna virus]   | 5.1     |
| 150    | Z25464                                | S.cerevisiae UNF1, LTV1, MRP8, CYB3 and TGL1 genes, complete CDS's              | 1.7     | 1255404  | (U53151) weak similarity to cytochrome b [Caenorhabditis elegans]   | 4.1     |
| 151    | U77846                                | Human elastin gene, partial cds and partial 3'UTR                               | 1.7     | 3355682  | (AL031124) putative secreted lyase  | 4.0     |
| 152    | X62880                                | S.scrofa mRNA for calcium release channel (CRC)                                 | 1.7     | 3327080  | (AB014533) KIAA0633 protein [Homo sapiens]  | 4.0     |
| 153    | Y00067                                | Human gene for neurofilament subunit M (NF-M)                                   | 1.7     | 479829   | heterogeneous ribonuclear particel protein homolog - Caenorhabditis elegans similarity to RNA recognition motifs [Caenorhabditis elegans]   | 3.9     |
| 154    | X68393                                | D.melanogaster gene for Beta-tubulin, exons 1 and 2                             | 1.7     | 2342682  | (AC000106) Contains similarity to Rattus AMP-activated protein kinase (gb X95577). [Arabidopsis thaliana]   | 3.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 155    | AB012284                              | Shuttle vector pAUR123 gene for Aurl-C, complete cds                            | 1.7     | 417704   | POL POLYPROTEIN (ORF1A/1B) [CONTAINS: RNA-DIRECTED RAN POLYMERASE; HELICASE; PROTEASE]  | 3.8     |
| 156    | M96633                                | Rattus norvegicus mitochondrial intermediate peptidase (MIP) mRNA, complete cds | 1.7     | 2314209  | (AE000163) H. pylori predicated coding region HP1054  | 3.1     |
| 157    | U49055                                | Rattus norvegicus CTD-binding SR-like protein rA8 mRNA complete cds             | 1.7     | 2497252  | INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 (IGFBP-4) (IBP-4) (IGF-BINDING PROTEIN 4) factor-binding protein-4 - sheep (fragment) factor-binding protein-4, IGFBP-4 [sheep, liver, Peptide, 237 aa] [Ovis aries] | 3.0     |
| 158    | Y15907                                | Mus musculus mRNA for myc-intron-binding protein-1                              | 1.7     | 912776   | iduronate-2-sulfatase, IDS {EC 3.1.6.13} Peptide Mutant, 550 aa]  | 3.0     |
| 159    | U67600                                | Methanococcus jannaschii section 142 of 150 of the complete genome              | 1.7     | 2982355  | (AF052252) fork head domain protein FKD9 [Danio rerio]  | 3.0     |
| 160    | AF013759                              | Homo sapiens calumein (Calu) mRNA, complete cds                                 | 1.7     | 2982355  | (AF052252) fork head domain protein FKD9 [Danio rerio]  | 2.9     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 161    | AF062915                              | Arabidopsis thaliana putative transcription factor (MYB90) mRNA, complete cds | 1.7     | 3878065  | (AL021066.1) Similarity to Human mRNA product KIAA0077 (TR:Q14997); cDNA EST yk243h8.5 comes from this gene; cDNA EST yk243h8.3 comes from this gene; cDNA EST yk359h4.5 comes from this gene [Caenorhabditis elegans] >gi 3880318 gnl PID e 349839 (Z81133) Similarity to Human mRNA product KIAA0077 (TR:Q14997); cDNA EST yk243h8.5 comes from this gene; cDNA EST yk243h8.3 comes from this gene; cDNA EST yk359h4.5 comes from this gene | 2.3     |
| 162    | X87526                                | H.sapiens genomic DNA (chromosome 3; clone NL3003R)                           | 1.7     | 3638957  | (AC004877) sco-spondin-mucin-like; similar to P98167 uncertain [Homo sapiens]   | 2.3     |
| 163    | AC005573                              | Homo sapiens chromosome 5, PAC clone 202e13                                   | 1.7     | 2465540  | (AF005632) phosphodiesterase I/nucleotide pyrophosphatase beta [Homo sapiens]   | 1.8     |
| 164    | D83402                                | Homo sapiens gene for prostacyclin synthase, exon 10 and complete cds         | 1.7     | 627608   | steroid hormone receptor TR3 - human sapiens]   | 1.7     |
| 165    | AF053700                              | Homo sapiens deltex (Dx) mRNA, complete cds                                   | 1.7     | 2662089  | (AB007864) KIAA0404 [Homo sapiens]  | 1.7     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 166                                   | AF043225       | Mus musculus 6-pyruvoyl-tetrahydropterin synthase (Pts) mRNA, complete cds   | 1.7     | 2352538   | (AF006564) alcohol dehydrogenase [Drosophila persimilis] persimilis]  | 1.4     |
| 167                                   | U52917         | Thermus aquaticus thermophilus NADH dehydrogenase I subunits NQO7 NQO6, NQO5, NQO4, NQO2, NQO1, NQO3, NQO8, NQO9, NQO10, NQO11, NQO12, NQO13, and NQO14, complete cds. | 1.7     | 2564334   | (AB006631) The human homolog of mouse Cux-2 [Homo sapiens]  | 1.0     |
| 168                                   | X72222         | M.musculus gene for serotonin 2 receptor   | 1.7     | 3875796   | (Z73425) Similarity to Yeast hypothetical YIK9 protein (SW:YIK9_YEAST); cDNA EST EMBL:T01252 comes from this gene; cDNA EST EMBL:D33205 comes from this gene; cDNA EST EMPL:D33955 comes from this gene; cDNA EST EMBL:D35484 co... | 1.0     |
| 169                                   | U23186         | Crotalus scutulatus PLA2-like pseudogene   | 1.7     | 853971  | (X83413) DR5 [Human herpesvirus 6] >gi 853972 (X83413) DR5 [Human herpesvirus 6]  | 0.99    |
| 170                                   | M83118         | Mus musculus factor VIII-associated protein (f8a) mRNA, complete cds.  | 1.7     | 3201617   | (AC004669) hypothetical protein [Arabidopsis thaliana]  | 0.80    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 171    | M38347                                | E.coli ATP-dependent proteinase (lon) gene, complete cds.                          | 1.7     | 4140322  | (AL031282) dJ283E3.3.2 (Cell Division Cycle-2-Like 2 (PITSLRE, p58/GTA, Galactosyltransferase Associated Protein Kinase)) (isoform beta 2-2) [Homo sapiens]   | 0.78    |
| 172    | U28838                                | Human transcription factor TFIIIB 90 kDa subunit                                   | 1.7     | 2495730  | HYPOTHETICAL PROLINE-RICH PROTEIN KIAA0269 >gi 1665805 gnl PID d1014089 (D87459) Similar to Volbox carteri extensin (S22697) [Homo sapiens]   | 0.62    |
| 173    | U72487                                | Rattus norvegicus calcium-independent alpha-latrotoxin receptor mRNA, complete cds | 1.7     | 544411   | GLYCOPROTEIN GP100 PRECURSOR (P29F8) discoideum]  | 0.35    |
| 174    | AE000718                              | Aquifex aeolicus section 50 of 109 of the complete genome                          | 1.7     | 2497569  | FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR (FGFR-3) (HEPARIN-BINDING GROWTH FACTOR RECEPTOR) >gi 2117851 pir  I5536 3 fibroblast growth factor receptor 3 - mouse >gi 199145 (M81342) fibroblast growth factor receptor 3 [Mus musculus] | 0.34    |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 175                                   | AF016897       | Oryza sativa GDP dissociation inhibitor protein OsGDI2 (OsGDI2) mRNA, complete cds   | 1.7     | 125362  | MACROPHAGE COLONY STIMULATING FACTOR I RECEPTOR PRECURSOR (CSF-1-R) (FMS PROTO-ONCOGENE) (C-FMS) factor 1 receptor - cat >gi 163855 (JO3149) M-CSF receptor [Felis domesticus] | 0.34    |
| 176                                   | U95102         | Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds                          | 1.7     | 85058   | muscarinic acetylcholine receptor - fruit fly acetylcholine receptor [Drosophila melanogaster]   | 0.20    |
| 177                                   | AF077352       | Chlamydomonas reinhardtii myosin heavy chain   | 1.7     | 728901  | ACROSOMAL PROTEIN SP-10 PRECURSOR SP-10 - western baboon >gi 2984818 bbs 127113 (S56458) SP-10=intracrosomal protein [Papio papio=baboons, Peptide, 285 aa] [Papio hymadryas]  | 0.20    |
| 178                                   | Z92788         | Caenorhabditis elegans cosmid F53B8, complete sequence [Caenorhabditis elegans]      | 1.7     | 746516  | (U23517) D1022.7 [Caenorhabditis elegans] >gi 3258651 elegans]   | 0.068   |
| 179                                   | AF002217       | Ralstonia eutropha megaplasmid pHG1 nitric oxide reductase (norB) gene, complete cds | 1.7     | 1143538   | (X87883) mitochondrial capsule selenoprotein [Rattus norvegicus] >gi 1354135 (U48702) mitochondria associated cysteine-rich protein SMCP                                       | 0.039   |



| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 180                                   | D30749         | Rat mRNA for protein tyrosine phosphatase   | 1.7     | 1228035   | (D83776) The KIAA0191 gene is expressed ubiquitously.; The KIAA0191 protein retains the C2H2 zinc-finger at its N-terminal region. [Homo sapiens] | 0.008   |
| 181                                   | M15202         | Rat fast skeletal TnT gene encoding troponin T isoforms, complete cds.  | 1.7     | 731172  | SKIN SECRETORY PROTEIN XP2 PRECURSOR  | 4e-04   |
| 182                                   | L07592         | Human peroxisome proliferator activated receptor mRNA, complete cds.  | 1.7     | 4033414   | PUTATIVE IMPORTIN BETA-4 SUBUNIT  | 2e-06   |
| 183                                   | U64031         | Dendrobium crumenatum ACC synthase gene, complete cds   | 1.7     | 3122885   | ASPARTYL-TRNA SYNTHETASE synthetase [Bacillus subtilis]   | 2e-11   |
| 184                                   | AF034970       | Homo sapiens docking protein (DOK-2) mRNA, complete cds   | 1.7     | 2289097   | (U78737) alpha(1,3)fucosyltransferase [Cricetulus griseus]  | 8e-12   |
| 185                                   | Z12839         | L.longiflorum mRNA encoding calmodulin. >:: gb L18912 LILCALM ODU Lilium longiflorum calmodulin mRNA, complete cds. | 1.7     | 2511747   | (AF023270) probable transcriptional regulator dre4  | 4e-12   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 186                                   | X53459         | Equine arteritis virus<br>(EAV) RNA genome<br>>::emb A45589 A45589<br>Sequence 1 from<br>Patent W09519438 >::<br>emb A58849 A58849<br>Sequence 1 from<br>Patent WO9700963<br>>::<br>gb AR013959 AR013959<br>Sequence 1 from<br>patent US 5773235 | 1.7     | 3979817   | (Z70683) Weak<br>similarity to Human<br>tyrosine-protein kinase<br>CSK<br>(SW:CSK_HUMAN);<br>cDNA EST<br>EMBL:C10908 comes<br>from this gene; cDNA<br>EST EMBL:C12822<br>comes from this gene;<br>cDNA EST yk408c2.3<br>comes from this gene;<br>cDNA EST yk408c2.5<br>... Human tyrosine-<br>protein kinase CSK<br>(SW:CSK_HUMAN);<br>cDNA EST<br>EMBL:C10908 comes<br>from this gene; cDNA<br>EST EMBL:C12822<br>comes from this gene;<br>cDNA EST yk408c2.3<br>come from this gene;<br>cDNA EST yk408c2.5<br>... | 1e-14   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 187    | K02668                                | E. coli ddl gene encoding D-alanine:D-alanine ligase and ftsQ and ftsA genes, complete cds, and ftsZ gene, 5' end. | 1.7     | 3879121  | (Z70310) predicted using Genefinder; Similarity to Mouse ankyrin (PIR Acc. No. S37771); cDNA EST EMBL:T01923 comes from this gene; cDNA EST EMBL:D32335 comes from this gene; cDNA EST EMBL:D32723 comes from this gene; cDNA ES... Genefinder; Similarity to Mouse ankyrin (PIR Acc. No. S37771); cDNA EST EMBL:T01923 comes from this gene; cDNA EST EMBL:D32335 comes from this gene; cDNA EST EMBL:D32723 comes from this gene; cDNA ES.... | 2e-19   |
| 188    | AB008375                              | Homo sapiens mRNA for osteoblast specific cystine-rich protein, complete cds                                       | 1.7     | 2496945  | HYPOTHETICAL 55.9 KD PROTEIN EEED8.6 IN CHROMOSOME II >gi 733603 (U23484) No definition line found [Caenorhabditis elegans]   | 1e-19   |
| 189    | L36603                                | Pseudomonas cepacia (clone Psudom70-1) heat shock protein 70 (hsp70) gene, complete cds                            | 1.7     | 2661842  | (Y15732) DNA polymerase beta [Xenopus laevis]   | 6e-20   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 190    | Z49760                                | P.blakesleeanus mRNA GTP cyclohydrolase I                                | 1.7     | 1731181  | HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II >gi 3874230 gnl PID e 1351618 protein (Swiss Prot accession number P38376); cDNA EST yk220e10.5 comes from this gene [Caenorhabditis elegans] | 3e-21   |
| 191    | U52428                                | Human fatty acid synthase gene, partial cds                              | 1.7     | 4226073  | (AF125443) contains similarity to S. pombe phosphatidyl synthase (GB:Z28295) [Caenorhabditis elegans]   | 6e-25   |
| 192    | U12767                                | Human mitogen induced nuclear orphan receptor                            | 1.6     | <NONE>   | <NONE>  | <NONE>  |
| 193    | Z63478                                | H.sapiens CpG DNA, clone 85a12, forward read cpg85a12.ft la.             | 1.6     | <NONE>   | <NONE>  | <NONE>  |
| 194    | AF084375                              | Homo sapiens inversin protein exons 8 and 9                              | 1.6     | <NONE>   | <NONE>  | <NONE>  |
| 195    | AE001114                              | Archaeoglobus fulgidus section 165 of 172 of the complete genome         | 1.6     | <NONE>   | <NONE>  | <NONE>  |
| 196    | AF084375                              | Homo sapiens inversin protein, exons 8 and 9                             | 1.6     | <NONE>   | <NONE>  | <NONE>  |
| 197    | U24217                                | Kluyveromyces lactis RNA polymerase II largest subunit gene, partial cds | 1.6     | <NONE>   | <NONE>  | <NONE>  |
| 198    | AE000580                              | Helicobacter pylori 26695 section 58 of 134 of the complete genome       | 1.6     | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 199    | X62083                                | H.sapiens mRNA for Drosophila female sterile homeotic (FSH) homologue<br>>::gb M80613 HUMFS HG Human homolog of Drosophila female sterile homeotic mRNA, complete cds. | 1.6     | <NONE>   | <NONE>   | <NONE>  |
| 200    | M28064                                | Plamodium brasilianum DNA homologous to the histidine-rich knob protein region of Plasmodium falcipraum.   | 1.6     | 457495   | (M26647) ORF X [Saccharomyces cerevisiae]                                    | 8.4     |
| 201    | U03114                                | Streptomyces albus lipase precursor (lip) gene, complete cds, and unidentified 5' ORF and 3' ORF, partial cds.   | 1.6     | 3638957  | (AC004877) sco-spondin-mucin-like;similar to P98167 uncertain [Homo sapiens] | 7.8     |
| 202    | U88422                                | Strix varia oocyte maturation factor Mos (c-mos) proto-oncogene, partial cds   | 1.6     | 137618   | VITAMIN D3 RECEPTOR (VDR) receptor [Rattus norvegicus]                       | 6.4     |
| 203    | M68519                                | Human pulmonary surfactant-associated protein SP-A (SFTP1) gene, complete cds.   | 1.6     | 3875423  | (Z38112) E03A3.6 [Caenorhabditis elegans]                                    | 4.9     |
| 204    | AF044575                              | Homo sapiens transcription factor POU4F3   | 1.6     | 2133625  | GABA transport protein-tobacco hornworm                                      | 4.7     |
| 205    | L48476                                | Homo sapiens (subcline 3_e10 from P1 H21) DNA sequence.  | 1.6     | 3687297  | (AJ005588) 5-epi-aristolochene synthase                                      | 4.6     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 206                                   | M18630         | Rat CNS 2',3'-cyclic nucleotide 3-phosphodiesterase                                  | 1.6     | 3880315   | (Z81133) Similarity to Human mRNA product KIAA0077 (TR:Q14997) [Caenorhabditis elegans]                  | 3.7     |
| 207                                   | AF027174       | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 1.6     | 267068  | TUMOR-ASSOCIATED ANTIGEN L6  | 3.6     |
| 208                                   | U53448         | Babesia microti heat shock protein 70 (hsp70) gene, complete cds                     | 1.6     | 1255429   | (U53155) strong similarity to the carboxyl two-thirds of valyl-tRNA synthetases [Caenorhabditis elegans] | 2.2     |
| 209                                   | AF084367       | Homo sapiens inversin protein mRNA, complete cds                                     | 1.6     | 1730076   | PROBABLE SERINE/THREONINE-PROTEIN KINASE CY49.28<br>>gi 1370255 gnl PID e 247094 (Z73966) pknJ           | 1.2     |
| 210                                   | D55635         | Yeast dis1 + gene for p93dis1, complete cds  | 1.6     | 3128353   | (AF010496) maltose transport inner membrane protein  | 1.2     |
| 211                                   | AF035756       | Streptomyces sp. 2-dehydro-3-deoxyphosphoheptona te aldolase gene, partial cds       | 1.6     | 853971  | (X83413) DR5 [Human herpesvirus 6] >gi 853972 (X83413) DR5 [Human herpesvirus 6]                         | 0.97    |
| 212                                   | X73479         | O.cuniculus rPTPA mRNA   | 1.6     | 3413810   | (Y17034) Bassoon [Mus musculus]  | 0.94    |
| 213                                   | X98330         | H.sapiens mRNA for ryanodine receptor 2  | 1.6     | 2072986   | (U95142) putative G-protein-coupled receptor G-protein-coupled receptor [Arabidopsis thaliana]           | 0.73    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 214    | X64194                                | P.anserina FMR1 gene exons 1 and 2  | 1.6     | 128014   | NECDIN<br>>gi 91129 pri  JN0148<br>necdin, brain - mouse<br>>gi 200020 (M80840)<br>necdin [Mus musculus]   | 0.42    |
| 215    | Z92788                                | Caenorhabditis elegans cosmid F53B8, complete sequence [Caenorhabditis elegans]                               | 1.6     | 746516   | (U23517) D1022.7 [Caenorhabditis elegans] >gi 3258651<br>elegans]  | 0.19    |
| 216    | AE000888                              | Methanobacterium thermoautotrophicum from bases 1098908 to 1112186 (section 94 of 148) of the complete genome | 1.6     | 462415   | INTERFERON-ALPHA/BETA RECEPTOR ALPHA CHAIN PRECURSOR (IFN-ALPHA-REC)<br>>gi 346520 pir  S2738<br>7 interferon alpha receptor type 1 - bovine >gi 432 | 0.001   |
| 217    | AB008375                              | Homo sapiens mRNA for osteoblast specific cysteine-rich protein, complete cds                                 | 1.6     | 2496945  | HYPOTHETICAL 55.9 KD PROTEIN EEED8.6 IN CHROMSOME II<br>>gi 733603 (U23484)<br>No definition line found [Caenorhabditis elegans]                     | 1e-18   |
| 218    | M25312                                | Orang-utan involucrin gene, complete cds.   | 1.6     | 3875131  | (Z70750) similar to vanadate resistance protein transmembranous domains [Caenorhabditis elegans]   | 3e-26   |
| 219    | AB012882                              | Cyprinus carpio mRNA for MyoD, complete cds   | 1.5     | <NONE>   | <NONE>   | <NONE>  |
| 220    | U29487                                | Caenorhabditis elegans cosmid C09C7   | 1.5     | <NONE>   | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |             |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-------------|---|---------|--|--|---------|
| SEQ ID                                | ACCES- SION | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 221                                   | X74760      | M.musculus mRNA for Notch 3   | 1.5     | 1364094  | integral membrane protein - Streptomyces pristinaespiralis >gi 872306 (X84072) integral membrane protein [Streptomyces pristinaespiralis]  | 4.3     |
| 222                                   | U72396      | Lycopersicon esculentum class II small heat shock protein Le-HSP17.6 mRNA, complete cds | 1.5     | 121855   | EXOGLUCANASE II PRECURSOR cellulose 1,4-beta cellobiosidase (EC 3.2.1.91) II precursor - fungus (Trichoderma reesei) 1,4-beta-cellobiosidase (EC 3.2.1.91) II - fungus cellobiohydrolase II [Trichoderma reesei] | 4.3     |
| 223                                   | U42391      | Human nyosin-IXb mRNA, complete cds   | 1.5     | 3688428  | (AJ011534) sucrose synthase  | 4.2     |
| 224                                   | M92296      | Pongo pygmaeus gamma-1 and gamma-2 globin genes, complete cds.                          | 1.5     | 186413   | (M13144) inhibin A [Homo sapiens]  | 0.22    |
| 225                                   | X94144      | C.japonica mRNA for QNR-71 protein  | 1.5     | 2745737  | (AF029791) UDP-Gal:betaGlcNac beta 1,3-galactosyltransferase-II [Mus musculus]   | 3e-08   |
| 226                                   | AB014557    | Homo sapiens mRNA for KIAA0657 protein, partial cds                                     | 1.5     | 1212992  | (X90568) Protein sequence and annotation available soon via Swiss-Prot; available at present via e-mail from LABEIT@EMBL-Heidelberg.DE [Homo sapiens]  | 4e-13   |
| 227                                   | AF000948    | Borrelia burgdorferi oligopeptide permease homolog OppAIV (oppAIV) gene, complete cds   | 1.3     | <NONE>   | <NONE>   | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 228    | AF057287                              | Mus musculus RAB/Rip protein mRNA, partial cds                                       | 1.3     | 2498005  | MYC PROTO-ONCOGENE PROTEIN (C-MYC) proto-oncogene [Sus scrofa]                                  | 2.6     |
| 229    | U38951                                | Drosophila melanogaster vacuolar ATPase subunit E                                    | 1.1     | <NONE>   | <NONE>  | <NONE>  |
| 230    | AF027148                              | Homo sapiens myogenic determining factor 3   | 1.1     | 3172134  | (U90209) RNA polymerase II largest subunit [Bonnemaisonia hamifera]                             | 2.3     |
| 231    | AF079310                              | Mus musculus histone deacetylase 3 (Hdac3) gene, exons 4 through 15 and complete cds | 1.0     | 1657601  | (U66220) unknown [Nannocystis exedens]  | 0.25    |
| 232    | X52134                                | P.radiata lac gene for laccase   | 0.95    | 996020   | (X91638) BRM protein [Gallus gallus]  | 0.31    |
| 233    | D89016                                | Human mRNA for Neuroblastoma, complete cds   | 0.93    | <NONE>   | <NONE>  | <NONE>  |
| 234    | X76392                                | C.familiaris VIP36 (vesicular integral-membrane protein of 36 kDa) mRNA              | 0.93    | 4176446  | (AL022238) dJ1042K10.2.1 (novel protein with probable rabGAP domains and Src homology domain 3) | 7e-81   |
| 235    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 0.90    | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 236    | AE000991                              | Archaeoglobus fulgidus section 116 of 172 of the complete genome   | 0.90    | 1176579  | EGT2 PROTEIN PRECURSOR (EARLY GT TRANSCRIPT 2) >gi 1362345 pri  S558 62 probable membrane protein YNL327w - yeast (Saccharomyces cerevisiae) cerevisiae] >gi 1302445 gnl PID e 239572 (Z71603) ORF YNL327w [Saccharomyces cerevisiae] | 6.9     |
| 237    | Z35922                                | S.cerevisiae chromosome II reading frame ORF YBR053c   | 0.86    | <NONE>   | <NONE>  | <NONE>  |
| 238    | U47331                                | Rattus norvegicus metabotropic glutamate receptor 4b mRNA, complete cds.   | 0.82    | 1550703  | (Z80225) hypothetical protein Rv2662  | 4.1     |
| 239    | X72810                                | H.sapiens Ig germline kappa-chain gene variable region (L3)  | 0.69    | 3023063  | (AF052587) F14 [Xylella fastidiosa]   | 6.7     |
| 240    | Z11700                                | Escherichia coli genes faeG, faeH, faeI, faeJ and IS629-like insertion sequence. >:: emb Z11710 ECFAEH IJ E.coli faeH, faeI and faeJ genes encoding FaeH, FaeI and FaeJ proteins | 0.69    | 2347188  | (AC002338) laccase isolog [Arabidopsis thaliana] thaliana]  | 3.9     |
| 241    | U71597                                | Phrynosoma douglassii NADH dehydrogenase subunit 4 (ND4) gene, mitochondrial gene encoding mitochondrial protein, partial cds  | 0.65    | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 242    | Z77798                                | Ammonia species LSU rRNA gene (partial; isolate Tr S 5; clone 16)         | 0.64    | 1174506  | GLUTAMYL-TRNA SYNTHETASE glutamate--tRNA ligase (EC 6.1.1.17) - Haemophilus influenzae (strain Rd KW20) >gi 1573240 (U32713) glutamyl-tRNA synthetase (gltX) [Haemophilus influenzae Rd]       | 1.2     |
| 243    | D25542                                | Human mRNA for golgi antigen gcp372, complete cds                         | 0.64    | 111230   | ultra-high-sulfur keratin 1 - mouse  | 1e-05   |
| 244    | M80234                                | Cow dopamine transporter mRNA, putative cds.                              | 0.64    | 3874972  | (Z99709) similar to Elongation factor Tu family (contains ATP/GTP binding P-loop); cDNA EST EMBL:D76223 comes from this gene; cDNA EST yk478c5.5 comes from this gene [Caenorhabditis elegans] | 8e-06   |
| 245    | AB007918                              | Homo sapiens mRNA for KIAA0449 protein, partial cds                       | 0.64    | 2833239  | EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >gi 530823 (U12535) epidermal growth factor receptor kinase substrate [Homo sapiens]  | 2e-14   |
| 246    | X51754                                | Human U266 rearranged DNA for lambda-immunoglobulin light chain           | 0.63    | 2072301  | (U95102) mitotic phosphoprotein 90 [Xenopus laevis]  | 1.5     |
| 247    | AE001554                              | Helicobacter pylori, strain J99 section 115 of 132 of the complete genome | 0.62    | <NONE>   | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 248                                   | Z64067         | H.sapiens CpG DNA,<br>cline 96e7, reverse<br>read cpg96e7.rta.   | 0.62    | <NONE>  | <NONE>  | <NONE>  |
| 249                                   | AJ223768       | Pinus sylvestris<br>microsatellite DNA,<br>clone SPAC11.5  | 0.62    | <NONE>  | <NONE>  | <NONE>  |
| 250                                   | AJ011592       | Bacteriophage P1 ban<br>gene   | 0.62    | 2493689   | PHOTOSYSTEM II<br>10 KD<br>PHOSPHOPROTEIN<br>[Populus deltoides]<br>>gi 2143326 gnl PID e<br>319090 (Y13328)<br>10kDa phosphoprotein<br>[Populus deltoides] | 7.9     |
| 251                                   | AF027151       | Xenopus laevis<br>survival of motor<br>neuron protein<br>interacting protein 1<br>(SIP1) mRNA,<br>complete cds | 0.62    | 4007790   | (AL034463) putative<br>single-strand<br>polynucleotide<br>binding protein<br>[Schizosaccharomyces<br>pombe]   | 2.0     |
| 252                                   | AJ000376       | Helobdella triserialis<br>mRNA for actin   | 0.62    | 1117968   | (U40763) CARS-Cyp<br>[Homo sapiens]<br>sapiens]   | 0.90    |
| 253                                   | M69231         | Rat thymosin beta 4<br>gene (pTB4G).intron.  | 0.62    | 4176370   | (AC005058) similar to<br>calcium-independent<br>phospholipase A2;<br>similar to AC004392<br>(PID:g3367519)<br>[Homo sapiens]                                | 6e-51   |
| 254                                   | AB021638       | Homo sapiens X11L2<br>mRNA for X11-like<br>protein 2, complete<br>cds  | 0.61    | <NONE>  | <NONE>  | <NONE>  |
| 255                                   | D26470         | Bacteriodes gingivalis<br>DNA for arginyl<br>endopeptidase,<br>complete cds                                    | 0.61    | <NONE>  | <NONE>  | <NONE>  |
| 256                                   | J04737         | A.thaliana ATPase<br>gene, complete cds.   | 0.61    | <NONE>  | <NONE>  | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 257                                   | U06756         | Bos taurus clone<br>bm1308 microsatellite<br>and are-1p repeat<br>region.   | 0.61    | 1922280   | (Y9905) snail like<br>protein [Gallus gallus]  | 0.51    |
| 258                                   | S75756         | p15=cylin D-<br>dependent kinases 4<br>and 6-binding<br>protein/p15 product<br>{exon/intron 1}<br>[human, brain tumors,<br>Genomic, 753 nt] | 0.61    | 484938  | hypothetical protein<br>253 - Streptomyces<br>griseus plasmid pSG1<br>(fragment)                                     | 0.13    |
| 259                                   | L39837         | Drosophila<br>melanogaster tumor<br>supressor (warts)<br>mRNA exons 1-8,<br>complete cds.   | 0.61    | 3875131   | (Z70750) similar to<br>vanadate resistance<br>protein<br>transmembranous<br>domains<br>[Caenorhabditis<br>elegans]   | 1e-09   |
| 260                                   | U52428         | Human fatty acid<br>synthase gene, partial<br>cds   | 0.61    | 4226073   | (AF125443) contains<br>similarity to S. pombe<br>phosphatidyl synthase<br>(GB:Z28295)<br>[Caenorhabditis<br>elegans] | 2e-26   |
| 261                                   | X15292         | Plasmodium<br>falciparum gene for<br>heat-shock protein<br>pPf203   | 0.60    | <NONE>  | <NONE>   | <NONE>  |
| 262                                   | AB020663       | Homo sapiens mRNA<br>for KIAA0856<br>protein, partial cds   | 0.60    | 470341  | (U00043) No<br>definition line found<br>[Caenorhabditis<br>elegans]  | 5.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 263    | U68723                                | Human checkpoint suppressor 1 mRNA, complete cds  | 0.60    | 544375   | GLACTOSE-BINDING PROTEIN REGULATOR<br>glucose/ galactose binding protein regulator - Agrobacterium tumefaciens<br>>gi 142228 (L10424)<br>glucose/ galactose binding protein regulator | 5.7     |
| 264    | M32687                                | S.griseus sporulation protein genes 1590 and 1422.  | 0.60    | 2582017  | (AF012871) Mergla' [Mus musculus]   | 3.3     |
| 265    | AJ005331                              | Homo sapiens NKCC2 gene, exon 4, isoform B  | 0.60    | 3128353  | (AF010496) maltose transport inner membrane protein   | 1.5     |
| 266    | U14103                                | Mus musculus RGL protein mRNA, complete cds.  | 0.60    | 4099845  | (U90533) serine protease inhibitor [Streptomyces fradiae]   | 0.098   |
| 267    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds   | 0.59    | 3282851  | (AF047897) ankyrin-like protein HGE-ANK [Ehrlichia sp. BDS]   | 5.5     |
| 268    | AE000872                              | Methanobacterium thermoautotrophicum from bases 896604 to 912784 (section 78 of 148) of the complete genome | 0.59    | 401553   | HYPOTHETICAL 24.5 KD PROTEIN IN NADB-SRMB INTERGENIC REGION   | 4.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 269    | L11871                                | Gallus gallus<br>achaetescute<br>homologue (ASH)<br>mRNA, complete cds.  | 0.59    | 628110   | Hypothetical protein -<br>human herpesvirus 4<br>reading frame 1<br>[Human herpesvirus<br>4] 2 [human<br>herpesvirus 4]<br>>gi 1334838 gnl PID e<br>25079 4 [Human<br>herpesvirus 4]<br>>gi 1334840 gnl PID e<br>25081 6 [Human<br>herpesvirus 4]<br>>gi 1334842 gnl PID e<br>25067 8 [Human<br>herpesvirus 4]<br>>gi 1334844 gnl PID e<br>25069 10 [Human<br>herpesvirus 4]<br>>gi 1334846 gnl PID e<br>25071 12 [Human<br>herpesvirus 4] | 4.2     |
| 270    | AF017114                              | Oryctolagus cuniculus<br>glycogen synthase<br>mRNA, complete cds   | 0.59    | 728856   | NITROGENASE<br>IRON-IRON<br>PROTEIN ALPHA<br>CHAIN<br>(NITROGENASE<br>COMPONENT I)<br>(DINITROGENASE)<br>capsulatus >gi 312238<br>(X70033) alternative<br>nitrogenase  | 2.4     |
| 271    | AF027807                              | Homo sapiens beta-<br>casein (CSN2) gene,<br>complete cds  | 0.59    | 3252932  | (AF067155) truncated<br>rev protein [Human<br>immunodeficiency<br>virus type 1]  | 1.5     |
| 272    | U81787                                | Human Wnt10B<br>mRNA, complete cds   | 0.59    | 3875538  | (Z67990) similar to<br>cuticle collagen  | 1.4     |
| 273    | U76036                                | Apteryx australis 16S<br>ribosomal RNA gene,<br>mitochondrial gene for<br>mitochondrial RNA,<br>partial sequence | 0.59    | 4193356  | (AF055088) ATP-<br>binding cassette; PsaB<br>[Streptococcus<br>pneumoniae]   | 0.83    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 274    | AB014564                              | Homo sapiens mRNA for KIAA0664 protein, partial cds                                      | 0.59    | 1709851  | PTB-ASSOCIATED SPLICING FACTOR (PSF) long form - human >gi 38458 (X70944) PTB-associated splicing factor [Homo sapiens] | 0.17    |
| 275    | AF044171                              | Homo sapiens cyclin-dependent kinase inhibitor 2D (CDKN2D) gene, partial cds             | 0.59    | 3925213  | (AL032626) Y37D8A.17 [Caenorhabditis elegans]   | 3e-10   |
| 276    | L19640                                | Saccharomyces cerevisiae cdc2/cdc28-related protein kinase gene, complete cds.           | 0.59    | 3880115  | (Z81130) T23G11.9 [Caenorhabditis elegans]  | 1e-21   |
| 277    | Z80999                                | Human DNA sequence from cosmid E140G5 on chromosome 22, complete sequence [Homo sapiens] | 0.58    | <NONE>   | <NONE>  | <NONE>  |
| 278    | Y11108                                | H.sapiens WNT8B gene   | 0.58    | <NONE>   | <NONE>  | <NONE>  |
| 279    | U80001                                | Sphyrana idiaestes lactate dehydrogenase A   | 0.58    | <NONE>   | <NONE>  | <NONE>  |
| 280    | Z49637                                | S.cerevisiae chromosome X reading frame ORF YJR137c                                      | 0.58    | <NONE>   | <NONE>  | <NONE>  |
| 281    | X64467                                | H.sapiens ALAD gene for porphobilinogen synthase   | 0.58    | <NONE>   | <NONE>  | <NONE>  |
| 282    | X74506                                | G.gallus hox B3 mRNA   | 0.58    | <NONE>   | <NONE>  | <NONE>  |
| 283    | U68040                                | Cochliobolus heterostrophus polyketide synthase  | 0.58    | <NONE>   | <NONE>  | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 284    | AF089084                              | Arbadiopsis thaliana putative auxin efflux carrier protein (PIN1) mRNA, complete cds   | 0.58    | <NONE>   | <NONE>   | <NONE>  |
| 285    | U38481                                | Rattus norvegicus ROK-alpha mRNA, complete cds   | 0.58    | <NONE>   | <NONE>   | <NONE>  |
| 286    | AF017656                              | Homo sapiens G protein beta 5 subunit mRNA, complete cds   | 0.58    | 3236249  | (AC004684) hypothetical protein [Arabidopsis thaliana]           | 9.2     |
| 287    | M96234                                | Human glutathione transferase class mu number 4  | 0.58    | 1280073  | (U55366) Similar to cuticle collagen [Caenorhabditis elegans]    | 7.1     |
| 288    | AB002339                              | Human mRNA for KIAA0341 gene, partial cds  | 0.58    | 861293   | (U28741) F35D2.1 gene product [Caenorhabditis elegans]           | 7.1     |
| 289    | U11295                                | Neisseria gonorrhoeae carbamoyl phosphate synthetase (glutamine) small subunit (carA) and large subunit (carB) genes, complete cds.  | 0.58    | 2425135  | (AF020283) DG2044 gene product [Dictyostelium discoideum]        | 5.3     |
| 290    | D80001                                | Human mRNA for KIAA0179 gene, partial cds  | 0.58    | 4097223  | (U49836) gamma-glutamyl transpeptidase precursor [Brugia malayi] | 4.1     |
| 291    | Z11700                                | Escherichia coli genes faeG, faeH, faeI, faeJ and IS629-like insertion sequence. >::emb Z11710 ECFA E HIJ E.coli faeH, faeI and faeJ genes encoding FaeH, FaeI and FaeJ proteins | 0.58    | 2347188  | (AC002338) laccase isolog [Arabidopsis thaliana] thaliana]       | 3.2     |
| 292    | M77350                                | Mouse hair keratin A1 (MKHA1) gene, complete cds.  | 0.58    | 141165   | HYPOTHETICAL 8.3 KD PROTEIN >gi 62179                            | 3.2     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 293                                   | X63787         | T.thermophila gene for snRNA U3-2   | 0.58    | 2826900   | (AB004461) DNA polymerase alpha catalytic subunit [Oryza sativa]  | 3.1     |
| 294                                   | D63881         | Human mRNA for KIAA0160 gene, partial cds   | 0.58    | 1934730   | (U95036) germin-like protein [Arabidopsis thaliana]   | 3.1     |
| 295                                   | U39378         | Gymnocarena mexicana 16S ribosomal RNA gene, mitochondrial gene encoding mitochondrial RNA, partial sequence        | 0.58    | 2194131   | (AC002062) Similar to Synechocystis antiviral protein   | 3.1     |
| 296                                   | X87987         | P.pastoris PRC1 gene >::dbj E12103 E12103 DNA encoding precursor of protease from Pichia pastoris                   | 0.58    | 3914197   | OCCLUDIN >gi 1276983 (U49221) occludin [Canis familiaris] >gfi 1589181 prf 2210347D occludin [Canis familiaris] | 3.1     |
| 297                                   | X75782         | A.thaliana (L.Heynh.) chloroplast mRNA for recombinant APS-kinase   | 0.58    | 1732444   | (D38529) DRPLA protein [Homo sapiens]   | 2.4     |
| 298                                   | M64848         | Mouse platelet-derived growth factor B chain musculus platelet-derived growth factor beta-chain (sis) gene, exon 5. | 0.58    | 3025832   | (AF055985) pyrrolidone-rich antigen [Onchocerca volvulus]   | 1.4     |
| 299                                   | AE001460       | Helicobacter pylori, strain J99 section 21 of 132 of the complete genome  | 0.58    | 2827198   | (AF037454) ubiquitin protein ligase [Mus musculus]  | 1.1     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 300                                   | X65720         | M.musculus gene for protein kinase C-gamma (exon 1 and exon 2)                            | 0.58    | 418395  | CHD1 PROTEIN<br>>gi 320737 pir  S3081<br>8 hypothetical protein<br>YER164w - yeast<br>(Saccharomyces cerevisiae) >gi 603404<br>(U18917) Chd1p:<br>transcriptional<br>regulator<br>[Saccharomyces cerevisiae] | 1.1     |
| 301                                   | AF043130       | Arabidopsis thaliana lactate dehydrogenase  | 0.58    | 3024637   | SEX-DETERMINING<br>REGION Y PROTEIN<br>determining protein<br>[Mus   | 0.62    |
| 302                                   | D28116         | Human genes for collagen type IV alpha 5 and 5, exon 1 and exon 1'                        | 0.58    | 1458250   | (U64835) T09D3.3<br>[Caenorhabditis elegans]   | 0.36    |
| 303                                   | AE001075       | Archaeoglobus fulgidus section 32 of 172 of the complete genome                           | 0.58    | 2276333   | (Z97991) hypothetical<br>protein Rv0336  | 0.36    |
| 304                                   | AF003948       | Rhodococcus opacus chloromuconate cycloisomerase transposase homolog genes, ccomplete cds | 0.58    | 477072  | mucin 7 precursor,<br>salivary - human   | 0.28    |
| 305                                   | U10692         | Human MAGE-7 antigen (MAGE7) pseudogene, complete cds.                                    | 0.58    | 3287858   | HOMEBOX<br>PROTEIN HOX-C11   | 0.054   |
| 306                                   | AF003948       | Rhodococcus opacus chloromuconate cycloisomerase transposase homolog genes, complete cds  | 0.58    | 3551821   | (AF058803) mucin 4<br>[Homo sapiens]   | 0.041   |
| 307                                   | X99350         | H.sapiens HFH4 gene, exon 1 and joined CDS  | 0.58    | 137483  | VAV PROTO-<br>ONCOGENE<br>>gi 55221 (X64361)<br>proto-oncogene (Mus musculus]  | 0.024   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 308    | AJ234282                              | Homo sapiens mRNA for Ig heavy chain variable region, clone C   | 0.58    | 3264846  | (AC003682) R27945_2 [Homo sapiens]                              | 0.018   |
| 309    | AF079310                              | Mus musculus histone deacetylase 3 (Hdac3) gene, exons 4 through 15 and complete cds  | 0.58    | 1657601  | (U66220) unknown [Nannocystis exedens]                          | 0.014   |
| 310    | AF019367                              | Human thiopurine methyltransferase (TPMT) gene, exons 6 and 7   | 0.58    | 3283352  | (AF063020) lens epithelium-derived growth factor [Homo sapiens] | 0.011   |
| 311    | X65720                                | M.musculus gene for protein kinase C-gamma (exon1 and exon 2)   | 0.58    | 1790878  | (U38291) microtubule-associated protein 1a [Homo sapiens]       | 0.008   |
| 312    | AB011155                              | Homo sapiens mRNA for KIAA0583 protein, partial cds   | 0.58    | 1351166  | SYNAPSINS IA AND IB >gi 163713                                  | 0.006   |
| 313    | X63692                                | H.sapiens mRNA for DNA  | 0.58    | 1817548  | (D84307) phosphoethanolamine cytidyltransferase [Homo sapiens]  | 0.001   |
| 314    | U53746                                | Feline immunodeficiency virus isolate FIV-Pco336-8 pol polyprotein (pol) gene, partial cds                                      | 0.58    | 2246532  | (U93872) ORF 73, contains large complex repeat CR 73            | 2e-05   |
| 315    | K00436                                | Rattus norvegicus (cline rt1-1) pseudo-Gly-tRNA gene.   | 0.58    | 206712   | (M64793) salivary proline-rich protein [Rattus norvegicus]      | 1e-05   |
| 316    | S79632                                | HSF2=heat shock factor 2 {alternatively spliced, splice junction region} [mice, CBA/J, testis, Genomic, 120 nt, segment 2 of 3) | 0.58    | 4038594  | (AJ222798) tDET1 protein [Lycopersicon esculentum]              | 3e-06   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 317    | D43964                                | Rat liver mRNA for Kan-1, complete cds                          | 0.58    | 1280135  | (U55376) coded for by C. elegans cDNA cm21e6; coded for by C. elegans cDNA cm01e2; similar to melibiose carrier protein (thiomethylgalactoside permease II) | 1e-08   |
| 318    | AB007918                              | Homo sapiens mRNA for KIAA0449 protein, partial cds             | 0.58    | 2833239  | EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >gi 530823 (U12535) epidermal growth factor receptor kinase substrate [Homo sapiens]                 | 3e-13   |
| 319    | AB001466                              | Homo sapiens mRNA for Efs1, complete cds                        | 0.58    | 2943716  | (D45027) 25 kDa trypsin inhibitor [Homo sapiens]  | 2e-14   |
| 320    | Z11701                                | Saccharomyces cerevisiae IRE1 gene for putative protein kinase. | 0.58    | 3880115  | (Z81130) T23G11.9 [Caenorhabditis elegans]  | 9e-21   |
| 321    | Z49535                                | S.cerevisiae chromosome X reading frame ORF YJR035w             | 0.58    | 4106562  | (Z83819) dJ146H21.2 (similar to CYTOCHROME B-245 HEAVY CHAIN) [Homo sapiens]  | 3e-33   |
| 322    | M62506                                | S.cerevisiae DBF20 gene, complete cds.                          | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 323    | X05944                                | Yeast PSS gene for phosphatidylserine synthetase                | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 324    | D38536                                | Snail gene for ADP-ribosyl cyclase, complete cds                | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 325    | Z75004                                | S.cerevisiae chromosome XV reading frame ORF YOR096w            | 0.57    | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 326    | L77034                                | Homo sapiens (subclone 10_e10 from P1 H16) DNA sequence.                                | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 327    | D37887                                | Cyprinus carpio c-myc gene for c-Myc, complete cds                                      | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 328    | AB014562                              | Homo sapiens mRNA for KIAA0662 protein, partial cds                                     | 0.57    | 197406   | (M57576) Ig kappa chain [Mus musculus]  | 8.9     |
| 329    | Z69651                                | Human DNA sequence from cosmid L75B9, Huntington's Disease Region, chromosome 4p16.3    | 0.57    | 1079280  | chaperonin containin TCP-1 complex gamma chain - African clawed frog >gi 793886 (X84990) Cctg                               | 8.9     |
| 330    | D89285                                | Mesocricetus auratus mRNA for inter-alpha trypsin inhibitor heavy chain 1, complete cds | 0.57    | 134132   | RYANODINE RECEPTOR, SKELETAL MUSCLE   | 6.9     |
| 331    | Z48951                                | S.cerevisiae chromosome XVI cosmid 9723   | 0.57    | 4210432  | (AJ130783) ACP2 protein [Mus musculus]  | 5.3     |
| 332    | X95573                                | A.thaliana mRNA for salt-tolerance zinc finger protein                                  | 0.57    | 1174828  | TYROSINE DECARBOXYLASE 2 4.1.1.25) - parsley >gi 169671 (M96070) tyrosine decarboxylase [Petroselinum]                      | 5.2     |
| 333    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                                 | 0.57    | 465646   | PROBABLE ABC TRANSPORTER ATP-BINDING PROTEIN IN NTRA/RPON 5'REGION (ORF1) Azorhizobium caulinodans >gi 311388 (X69959) ORF1 | 4.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 334    | AE001116                              | <i>Borrelia burgdorferi</i> (section 2 of 70) of the complete genome                                 | 0.57    | 2314735  | (AE000653) Na <sup>+</sup> /H <sup>+</sup> antiporter (nhaA) [ <i>Helicobacter pylori</i> 26695]   | 4.0     |
| 335    | Z34291                                | <i>R.norvegicus</i> mRNA for putative chloride channel.  | 0.57    | 1350832  | DNA-DIRECTED RNA POLYMERASE I SECOND LARGEST SUBUNIT (RNA POLYMERASE I SUBUNIT 2) chain RPA2 - <i>Euplotes octocarinatus</i> (SGC9) >gi 578407 octocarinatus]  | 3.0     |
| 336    | D88255                                | <i>Homo sapiens</i> A30 Vk germline gene, partial cds  | 0.57    | 3875983  | (Z81063) similar to Actinin-type actin-binding domain containing proteins [ <i>Caenorhabditis elegans</i> ]  | 3.0     |
| 337    | AF037261                              | <i>Homo sapiens</i> SH3-containing adapter molecule-1 mRNA, complete cds                             | 0.57    | 1397341  | (U61955) Similar to kinesin-like protein; coded for by <i>C. elegans</i> cDNA yk184h5.3; coded for by <i>C. elegans</i> cDNA yk184h5.5; coded for by <i>C. elegans</i> cDNA yk13d7.3; coded for by <i>C. elegans</i> cDNA yk13d7.5; coded for by <i>C. elegans</i> cDNA yk31e1.5; co... >gi 3493541 (AF057567) kinesin-like protein ZEN-4a [ <i>Caenorhabditis elegans</i> ] | 2.3     |
| 338    | U26595                                | <i>Rattus norvegicus</i> prostaglandin F2a receptor regulatory protein precursor, mRNA, complete cds | 0.57    | 2773160  | (AF039656) neuronal tissue-enriched acidic protein [ <i>Homo sapiens</i> ]   | 2.3     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 339                                   | X69903         | R.norvegicus mRNA for interleukin-4 receptor  | 0.57    | 2649193   | (AE)001009) quinone-reactive Ni/Fe-hydrogenase B-type cytochrome subunit (hydC) [Archaeoglobus fulgidus]  | 1.8     |
| 340                                   | Z74825         | S.cerevisiae chromsome XV reading fram ORF YOL083w  | 0.57    | 1458319   | (U64846) F47D2.5 gene product [Caenorhabditis elegans]  | 1.4     |
| 341                                   | AJ131469       | Foot-and-mouth disease virus O vpl gene, strain O/A/58  | 0.57    | 91206   | proline-rich protein - mouse (fragment) musculus]   | 1.4     |
| 342                                   | AF011360       | Mus musculus regulator of G-protein signaling 7 (RGS7) mRNA, complete cds   | 0.57    | 542514  | gelsolin - American lobster   | 0.80    |
| 343                                   | AF011360       | Mus musculus regulator of G-protein signaling 7 (RGS7) mRNA, complete cds   | 0.57    | 1078946   | gelsolin - American lobster >gi 452313 gelsolin [Homarus americanus]  | 0.80    |
| 344                                   | L39210         | Homo sapiens inosine monophosphate dehydrogenase type II gene, complete cds   | 0.57    | 559526  | (X77466) 98.8kD polypeptide [Strawberry latent ringspot virus]  | 0.79    |
| 345                                   | U81523         | Human endometrial bleeding associated factor mRNA, complete cds   | 0.57    | 211499  | (K01702) HMW/LMW collagen subunit precursor [Gallus gallus]   | 0.79    |
| 346                                   | U46561         | Tetrahymena thermophila polyubiquitin (TTU3) gene, complete cds, and RNA polymerase II subunit 2 (RPB2) gene, partial cds | 0.57    | 2506493   | HYPOTHETICAL 100.5 KD PROTEIN IN IAP-CYSH INTERGENIC REGION >gi 882654 (U29579) alternate gene name ygcB; ORF_f888 [Escherichia coli] >gi 1789119 | 0.60    |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 347    | X95543                                | C.japonica mRNA for legumin (clone CjLeg31)                 | 0.57    | 1709261  | NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M) >gi 1083164 pir  S553 95 neurofilament protein M - rabbit (fragment) >gi 854353 | 0.46    |
| 348    | Y17282                                | Homo sapiens mRNA for cytokeratin type II                   | 0.57    | 3044086  | (AF055904) unknown [Myxococcus xanthus]   | 0.45    |
| 349    | X00716                                | Frog mRNA fragment for alpha-A2-crystallin                  | 0.57    | 3406654  | (AF079369) transcriptional repressor TUP1 [Dictyostelium discoideum]  | 0.20    |
| 350    | X53238                                | Klebsiella sp. bacteriophage K11 gene 1 for RNA polymerase  | 0.57    | 1228093  | (Z46913) polyketide synthase  | 0.16    |
| 351    | X99012                                | H.spaiens FUS gene, exon 12                                 | 0.57    | 243898   | (S78897) GOR=antegenic epitope [chimpanzees, Peptide, 427 aa] [Pan]   | 0.090   |
| 352    | A1008711                              | Human DNA sequence from PAC 390N22 on chromosome Xp22.2     | 0.57    | 1469545  | (U53585) fibronectin attachment protein [Mycobacterium avium]   | 0.053   |
| 353    | S74506                                | SOX9 [human, fetal brain, Genomic, 1494 nt. segment 3 of 3] | 0.57    | 1326350  | (U58748) similar to potential transmembrane domains in S. cerevisiae nuclear division RFT1 protein (SP:P38206)  | 0.017   |
| 354    | D25542                                | Human mRNA for golgi antigen gcp372, complete cds           | 0.57    | 4063399  | (AF102575) cell surface protein DTFA [Dictyostelium discoideum]   | 0.005   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 355                                   | AB015426       | Mus musculus mRNA for alpha 1,3-fucosyltransferase IX, complete cds  | 0.57    | 2661842   | (Y15732) DNA polymerase beta [Xenopus laevis]   | 7e-11   |
| 356                                   | X51394         | Xenopus mRNA for APEG protein, containing a highly repetitive amino acid sequence  | 0.57    | 1929056   | (Y12090) putative 3,4-dihydroxy-2-butanone kinase [Lycopersicon esculentum]   | 9e-12   |
| 357                                   | AB007918       | Homo sapiens mRNA for KIAA0449 protein, partial cds  | 0.57    | 2833239   | EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >gi 530823 (U12535) epidermal growth factor receptor kinase substrate [Homo sapiens] | 3e-13   |
| 358                                   | AB001466       | Homo sapiens mRNA for Efs1, complete cds   | 0.57    | 2943716   | (D45027) 25 kDa trypsin inhibitor [Homo sapiens]  | 2e-14   |
| 359                                   | Y00760         | Rabbit mRNA for adult fast skeletal troponin-C   | 0.57    | 2576348   | (AC002400) Glutamyl tRNA synthetase [Homo sapiens]  | 2e-28   |
| 360                                   | X95153         | H.sapiens brca2 gene exon 3 >:: emb A62778 A62778 Sequence 19 from Patent WO9719110  | 0.57    | 3419847   | (AC004982) similar to yeast hypothetical protein ybk4; similar to P38164 (PID:g586461) [Homo sapiens]                                       | 2e-55   |
| 361                                   | X85967         | B.vulgaris mRNA for betavulgin   | 0.56    | <NONE>  | <NONE>  | <NONE>  |
| 362                                   | U09251         | Mycoplasma genitalium DNA gyrase subunit B complete cds, DNA polymerase III beta subunit (dnaN) and seryl-tRNA synthetase (serS) genes, partial cds. | 0.56    | <NONE>  | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION | P VALUE |
| 363    | V00158                                | Chloroplast <i>Euglena gracilis</i> genes coding for transfer RNAs specific for threonine, glycine, methionine, serine and glutamine.      | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 364    | D88151                                | <i>Clostridium perfringens</i> DNA for D-alanine:D-alanine ligase, cortical fragment-lytic enzyme  | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 365    | U67478                                | <i>Methanococcus jannaschii</i> section 20 of 150 of the complete genome   | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 366    | L23800                                | <i>Tachyglossus aculeatus</i> beta-globin homolog (HBB) gene, complete cds   | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 367    | AB011129                              | Homo sapiens mRNA for KIAA0557 protein, partial cds  | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 368    | L77034                                | Homo sapiens (subclone 10_e10 from P1 H16) DNA sequence.   | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 369    | Z47202                                | <i>C.albicans</i> gene for TFIIB (BRF1) subunit.   | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 370    | U53868                                | <i>Clostridium acetobutylicum</i> mannitol-specific phosphotransferase system (PTS) system, mtlA, mtlR, mtlF, and mtlD genes, complete cds | 0.56    | <NONE>   | <NONE>      | <NONE>  |
| 371    | AF041259                              | Homo sapiens breast cancer putative transcription factor (ZABC1) mRNA, complete cds  | 0.56    | <NONE>   | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 372    | L42636                                | Plasmodium falciparum variant-specific surface protein (var-7) mRNA, complete cds. | 0.56    | 2213557  | (Z97052) hypothetical protein   | 8.8     |
| 373    | U96180                                | Human protein tyrosine phosphatase (TEP1) mRNA, complete cds                       | 0.56    | 731016   | THIOREDOXIN REDUCTASE thioredoxin reductate (NADPH) [Coxiella burnetii]   | 8.7     |
| 374    | L76259                                | Homo sapiens PTS gene, complete cds  | 0.56    | 2369863  | (Y12225) Spi-1/PU.1 transcription factor  | 6.7     |
| 375    | AF045946                              | Mus musculus D16Jhu17 YAC 98B3 acentric end, partial sequence                      | 0.56    | 2130017  | hypothetical protein - common sunflower protein [Helianthus annuus]   | 5.1     |
| 376    | X97986                                | M.musculus mRNA for desmocollin type I   | 0.56    | 4038031  | (AC005936) hypothetical protein [Arabidopsis thaliana]  | 3.9     |
| 377    | X79437                                | M.musculus whey acidic protein (WAP) gene, exon 1                                  | 0.56    | 549670   | SPINDLE POLE BODY COMPONENT SPC42 yeast (Saccharomyces cerevisiae) >gi 486054 (Z28042) ORF YKL042w [Saccharomyces cerevisiae] >gi 666098 (X71621) hypothetical 42.3 kD protein [Saccharomyces cerevisiae] | 3.9     |
| 378    | M27902                                | Rat cardiac specific sodium channel alpha-subunit mRNA, complete cds.              | 0.56    | 585234   | ENDOGLUCANASE G PRECURSOR 3.2.1.-) CelCCG precursor - Clostridium cellulolyticum cellulolyticum]  | 3.9     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 379    | AF036696                              | Caenorhabditis elegans cosmid F15B10  | 0.56    | 546071   | gp70=enevelope protein {endogenous provirus} host=cat lymphoid tissues, Peptide, 445 aa]  | 3.6     |
| 380    | Z99102                                | Caenorhabditis elegans cosmid B0331, complete sequence [Caenorhabditis elegans] | 0.56    | 603664   | (U14101) putative reverse transcriptase; ORF2; encodes aa motifs conserved in reverse transcriptases; most closely related reverse transcriptases are those of non-LTR retrotransposons. The 3'901 bp of this CDS are identical of the 3'901 bp ... | 3.0     |
| 381    | L27850                                | Equus caballus (clone T131) T-cell receptor DNA, V-region.                      | 0.56    | 1079150  | transcription factor shn - fruit fly  | 1.7     |
| 382    | X97986                                | M.musculus mRNA for desmocollin type 1  | 0.56    | 2497227  | HYPOTHETICAL 113.1 KD PROTEIN IN PRE5-FET4 INTERGENIC REGION >gi 1072409 (Z54141) unknown   | 1.7     |
| 383    | AF087455                              | Didelphis virginiana G protein receptor kinase 2 mRNA, complete cds             | 0.56    | 1213453  | (U12964) contains ankyrin-like repeats; similar to human desmoplakin repeat region [Caenorhabditis elegans]   | 1.3     |
| 384    | D80011                                | Human mRNA for KIAA0189 gene, complete cds                                      | 0.56    | 226535   | protease [Hepatitis B virus]  | 1.1     |
| 385    | AJ002272                              | Mus musculus mRNA for HAP1-A protein, 3' region                                 | 0.56    | 3327158  | (AB014572) KIAA0672 protein [Homo sapiens]  | 1.0     |
| 386    | L39210                                | Homo sapiens inosine monophosphate dehydrogenase type II gene, complete cds     | 0.56    | 628431   | coat protein - strawberry latent ringspot virus   | 0.77    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 387    | X02770                                | Mouse Thy-1.2 gene 5' untranslated region and exon 1  | 0.56    | 3327046  | (AB014516) KIAA0616 protein [Homo sapiens]  | 0.59    |
| 388    | AF038575                              | Schizosaccharomyces pombe Wiskott-Aldrich Syndrome protein homolog (wsp1+) gene, complete cds, and BTF3/beta-NAC gene, partial sequence | 0.56    | 88466  | salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human >gi 190484 (K03203) prepro salivary proline-rich protein [Homo sapiens] >gi 190512 | 0.35    |
| 389    | X56747                                | Rat mRNA for fetal intestinal lactase-phlorizin hydrolase precursor, partial  | 0.56    | 2072742  | (Z48674) chitinase homologue [Sesbania rostrata]  | 0.23    |
| 390    | Y12072                                | G.arboreum mRNA for farnesyl pyrophosphate synthase   | 0.56    | 296670   | (X07882) Po protein [Homo sapiens]  | 0.20    |
| 391    | S75756                                | p15=cyclin D-dependent kinases 4 and 6-binding protein/p15 product {exon/intron 1} [human, brain tumors, Genomic, 753 nt]               | 0.56    | 1082743  | protein kinase (EC 2.7.1.37) SPRK - human sapiens] >gi 1090771 prf 2019 37A protein Tyr kinase I  | 0.15    |
| 392    | U62528                                | Equus caballus type II collagen mRNA, complete cds  | 0.56    | 461671   | [Segment 1 of 2] COLLAGEN ALPHA 1(I) CHAIN  | 0.030   |
| 393    | X96877                                | C.reinhardtii mRNA for unknown lumenal polypeptide  | 0.56    | 3341678  | (AC003672) putative zinc finger protein [Arabidopsis thaliana]  | 5e-09   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 394    | S78788                                | cGATA-3 [chickens, liver, Genomic, 979 nt, segment 4 of 4]           | 0.56    | 2661590  | (AL009196) 1-evidence=predicated by content; 1-method=genefinder;840; 1-method_score=59.41; 1-evidence_end; 2-evidence=predicated by match; 2-match_accession=AA950019; 2-match_description=LD29959.5p rime LD Drosophila melanogas... | 2e-11   |
| 395    | AF006640                              | Drosophila melanogaster Ste20-like protein kinase mRNA, complete cds | 0.56    | 1109830  | (U41534) coded for by C. elegans cDNA CEESI42F; Similar to helicases of SNF2/RAD54 family. [Caenorhabditis elegans]  | 6e-12   |
| 396    | AF006640                              | Drosophila melanogaster Ste20-like protein kinase mRNA, complete cds | 0.56    | 1109830  | (U41534) coded for by C. elegans cDNA CEESI42F; Similar to helicases of SNF2/RAD54 family. [Caenorhabditis elegans]  | 4e-13   |
| 397    | AE000716                              | Aquifex aeolicus section 48 of 109 of the complete genome            | 0.56    | 3688350  | (AL030996) dJ1189B24.4 (novel PUTATIVE protein similar to hypothetical proteins S. pombe C22F3.14C and C. elegans C16A3.8) [Homo sapiens]  | 3e-66   |
| 398    | Z36079                                | S.cerevisiae chromosome II reading frame ORF YBR210w                 | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 399    | Y17267                                | Mus musculus mRNA for ubiquitin conjugating enzyme                   | 0.55    | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 400    | AC001461                              | Homo sapiens (subclone 2_g5 from BAC H107) DNA sequence                                | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 401    | AF019079                              | Alouatta seniculus breast and ovarian susceptibility (BRCA1) gene, partial cds         | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 402    | M90058                                | Human serglycin gene, exons 1,2, and 3.  | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 403    | AB013469                              | Mus musculus CLM2 gene for cytohesin 2, complete and partial cds, alternative splicing | 0.55    | 1729760  | (Z68152) chitinase [Gossypium hirsutum]  | 8.6     |
| 404    | AJ011592                              | Bacteriophage P1 ban gene  | 0.55    | 2493689  | PHOTOSYSTEM II 10 KD PHOSPHOPROTEIN [Populus deltoides] >gi 2143326 gnl PID e 319090 (Y13328) 10kDa phosphoprotein [Populus deltoides] | 6.6     |
| 405    | Z15118                                | T.brucei kinetoplast maxicircle variable region DNA                                    | 0.55    | 2970432  | (AF049132) NADH dehydrogenase subunit 5 [Florometra serratissima]  | 6.5     |
| 406    | Z48951                                | S.cerevisiae chromosome XVI cosmid 9723  | 0.55    | 4210432  | (AJ130783) APC2 protein [Mus musculus]   | 4.9     |
| 407    | U78726                                | Homo sapiens mad protein homolog Smad2 gene, promoter, exon 1a and exon 1b             | 0.55    | 3319290  | (AF055994) thyroid hormone receptor-associated protein complex component TRAP220 [Homo sapiens]  | 4.9     |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 408                                   | AG001389       | Homo sapiens<br>genomic DNA, 21q<br>region, clone:<br>9H11Bm42   | 0.55    | 125684  | KRUEPPEL<br>PROTEIN<br>>gi 72899 pir  TWFF<br>Krueppel gap protein -<br>fruit fly (Drosophila<br>sp.) melanogaster]<br>>gi 224875 prf  12023<br>48A Krueppel gene | 3.8     |
| 409                                   | M27640         | Plasmodium vivax<br>major blood stage<br>surface antigen gene,<br>partial cds.                                 | 0.55    | 549453  | X-LINKED PEST-<br>CONTAINING<br>TRANSPORTER<br>transporter - human<br>>gi 458255 (U05321)<br>X-linked PEST-<br>containing transporter<br>[Homo sapiens]           | 3.8     |
| 410                                   | D37977         | Fugu rubripes mRNA<br>for sodium channel<br>alpha subunit, partial<br>cds                                      | 0.55    | 1435038   | (D38024) ORF [Homo<br>sapiens]  | 3.7     |
| 411                                   | M88505         | Ostertagia ostertagi<br>cathepsin B-like<br>cysteine protease<br>gene, partial cds.                            | 0.55    | 3941277   | (AF000900) p45<br>[Rattus norvegicus]   | 2.9     |
| 412                                   | U95098         | Xenopus laevis<br>mitotic<br>phosphoprotein 44<br>mRNA, partial cds  | 0.55    | 2570154   | (AB008376) 17-kDa<br>PKC-potentiased<br>inhibitory protein of<br>PP1 [Sus scrofa]   | 2.8     |
| 413                                   | U89241         | Human mibp gene,<br>partial cds  | 0.55    | 4097465   | (U62253) 16kDa<br>secretory protein [Sus<br>scrofa]   | 2.2     |
| 414                                   | AF027151       | Xenopus laevis<br>survival of motor<br>neuron protein<br>interacting protein 1<br>(SIP1) mRNA,<br>complete cds | 0.55    | 4007790   | (AL034463) putative<br>single-strand<br>polynucleotide<br>binding protein<br>[Schizosaccharomyces<br>pombe]   | 1.7     |
| 415                                   | AF006821       | Burfo marinus<br>natiuretic peptide<br>receptor C mRNA,<br>partial cds   | 0.55    | 2245075   | (Z97343) GTP-<br>binding RAB2A<br>protein   | 1.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 416    | Y12736                                | Lactococcus lactis cremoris plasmid pJW565 DNA, lllabiiM, lllabiiR genes and orfX                    | 0.55    | 3386334  | (AF035120) type I procollagen pro-alpha 2 chain [Canis familiaris] | 1.3     |
| 417    | U38307                                | Mus musculus collagen alpha-1 type 1 gene, 5' flanking region, partial sequence.                     | 0.55    | 1362802  | gastric mucin - human (fragment) >gi 547517                        | 1.3     |
| 418    | D13473                                | Mouse mRNA for Rad51 protein   | 0.55    | 1374698  | (D83032) nuclear protein, NP220 [Homo sapiens]                     | 1.3     |
| 419    | AF045238                              | Bungarus fasciatus acetylcholinesterase gene, alternatively spliced products, partial cds            | 0.55    | 3261734  | (Z94752) hypothetical protein Rv1004c                              | 0.99    |
| 420    | AE000795                              | Methanobacterium thermoautotrophicum from bases 1 to 10208 (section 1 of 148) of the complete genome | 0.55    | 186396   | (M94131) mucin [Homo sapiens]                                      | 0.97    |
| 421    | X99537                                | Y.lipolytica SEC62 gene  | 0.55    | 3876397  | (Z81068) F25H5.2 [Caenorhabditis elegans]                          | 0.58    |
| 422    | U08147                                | Aquilegia sp. phytochrome (PHYB/D) gene, partial cds.  | 0.55    | 2338024  | (AF005370) ribonucleotide-reductase, large subunit                 | 0.57    |
| 423    | Z56586                                | H.sapiens CpG DNA, clone 12c8, reverse read cpg12c8.rtld.  | 0.55    | 3320122  | (U46007) espin [Rattus norvegicus]                                 | 0.44    |
| 424    | U39442                                | Mus musculus glutamine:fructose-6-phosphate amidotransferase (GFAT) gene, 5' region and partial cds  | 0.55    | 282600   | hypothetical protein - Mycoplasma hyorhinis                        | 0.43    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 425    | K02298                                | Rat chymotrypsin B (chyB) gene, complete cds.   | 0.55    | 3413810  | (Y17034) Bassoon [Mus musculus]  | 0.33    |
| 426    | X84792                                | M.musculus clusterin gene   | 0.55    | 1652475  | (D90905) hypothetical protein  | 0.25    |
| 427    | U00185                                | Capra aegagrus Saanen and Weisse Edel breeds DR beta-chain antigen binding domain, MHC class II DRB | 0.55    | 2507136  | SUBTILIN BIOSYNTHESIS PROTEIN SPAB   | 0.19    |
| 428    | Z54946                                | H.sapiens CpG DNA, clone 178a12, reverse read cpg178a12.rtla  | 0.55    | 807646   | (M17294) unknown protein [Human herpesvirus 4]   | 0.065   |
| 429    | AF031650                              | Oryctolagus cuniculus anion exchanger 3 brain isoform (AE3) mRNA, complete cds                      | 0.55    | 1778210  | (U68412) fibrillar collagen [Arenicola marina]   | 0.044   |
| 430    | m25579                                | Bovine adenylyl cyclase Type I mRNA, complete cds.  | 0.55    | 2649040  | (AE000997) conserved hypothetical protein [Archaeoglobus fulgidus]   | 0.023   |
| 431    | Z48796                                | H.sapiens Ski-W mRNA for helicase   | 0.55    | 330452   | (M14708) DNA polymerase [Human cytomegalovirus]  | 0.023   |
| 432    | M80234                                | Cow dopamine transporter mRNA, putative cds.  | 0.55    | 3874972  | (Z99709) similar to Elongation factor Tu family (contains ATP/GTP binding P-loop); cDNA EST EMBL:D76223 comes from this gene; cDNA EST yk478c5.5 comes from this gene [Caenorhabditis elegans] | 4e-04   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 433                                   | U91616         | Human I kappa B epsilon (IkBe) mRNA, complete cds                                 | 0.55    | 3875577   | (Z68314) similar to G-protein; cDNA EST EMBL:C11959 comes from this gene; cDNA EST EMBL:C10341 comes from this gene; cDNA EST yk494e4.3 comes from this gene; cDNA EST yk448a8.5 comes from this gene; cDNA EST EMBL:C10341 comes from this gene; cDNA EST yk494e4.3 comes from this gene; cDNA EST yk448a8.5 comes from this gene [Caenorhabditis elegans]<br>>gi 3880364 gnl PID e1349948 (Z83016) similar to G-protein; cDNA EST EMBL:C11959 comes from this gene; cDNA EST EMBL:C10341 comes from this gene; cDNA EST yk494e4.3 comes from this gene; cDNA EST yk448a8.5 comes from this gene [Caenorhabditis elegans] | 7e-06   |
| 434                                   | D10910         | Arabidopsis thaliana Atpk7 gene for serine/threonine protein kinase, complete cds | 0.55    | 3876072   | (Z81505) Similarity to Metanococcus hypothetical protein 0682 (TR:Q58095) [Caenorhabditis elegans]   | 4e-42   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |                                   |         |
|--------|---------------------------------------|---|---------|--|-----------------------------------|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION                       | P VALUE |
| 435    | L22013                                | Swinepox virus complete ORFS C20L-C1L >:: gb I58297 I58297 Sequence 14 from patent US 5651972   | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 436    | Z92653                                | Human immunodeficiency virus type 1 env gene  | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 437    | K01992                                | E.coli phosphate-repressible periplasmic phosphate-binding protein (phoS), peripheral membrane proteins (pstC, pstB and phoU) and integral membrane protein (pstA) genes, complete cds. | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 438    | AE001415                              | Plasmodium falciparum chromosome 2, section 52 of 73 of the complete sequence   | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 439    | AF064030                              | Helianthus tuberosus lectin 2 mRNA, complete cds  | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 440    | X12591                                | E.coli plasmid DNA for colicin E9   | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 441    | U73679                                | Caenorhabditis elegans YNK1-a mRNA, complete cds  | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 442    | Z93990                                | Unidentified bacterium DNA for 16S ribosomal RNA  | 0.54    | <NONE>   | <NONE>                            | <NONE>  |
| 443    | X85967                                | B.vulgaris mRNA for betavulgin  | 0.54    | 757836   | (Z37980) ORF12 [Escherichia coli] | 8.3     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 444                                   | U76524         | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                          | 0.54    | 151377  | (M80653) tetraheme [Pseudomonas stutzeri]             | 6.2     |
| 445                                   | X71800         | H.sapiens gene for 5S rRNA (640 bp) >:: emb[X71801]HS5SR6 40B H.sapiens gene for 5S rRNA (640 bp)  | 0.54    | 3322653   | (AE001216) T. pallidum predicted coding region TP0369 | 2.7     |
| 446                                   | U89241         | Human mibp gene, partial cds   | 0.54    | 4097465   | (U62253) 16kDa secretory protein [Sus scrofa]         | 2.2     |
| 447                                   | L16013         | Rattus norvegicus Q-like gene sequence   | 0.54    | 3087760   | (AJ005583) p75 protein [Cryptocodium cohnii]          | 0.95    |
| 448                                   | U60275         | Capra hircus skeletal muscle voltage-gated chloride channel gCIC-1 mRNA, partial cds               | 0.54    | 1781344   | (Y10438) FK506 polyketide synthase                    | 0.95    |
| 449                                   | U36795         | Myxococcus xanthus rfbABC O-antigen biosynthesis operon, rfbA, rfbB, and rfbC genes, complete cds. | 0.54    | 3877232   | (Z81540) predicted using Genefinder                   | 0.74    |
| 450                                   | AF053091       | Drosophila melanogaster eyelid (eld) mRNA, complete cds  | 0.54    | 2144110   | zinc finger protein RIZ - rat >gi 949996              | 0.14    |
| 451                                   | V00602         | Genome of the bacteriophage fd (Inoviridae).   | 0.54    | 2661620   | (AL009197) hypothetical protein                       | 0.11    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 452    | U60800                                | Human semaphorin (CD100) mRNA, complete cds   | 0.54    | 125682   | KERATIN, ULTRA HIGH-SULFUR MATRIX PROTEIN (UHS KERATIN)<br>>gi 109116 pir  A36686 ultra-high-sulfur keratin - sheep<br>>gi 1306 (X55294) ultra high-sulphur keratin protein [Ovis aries]       | 0.003   |
| 453    | X85969                                | S.coelicolor secD, secF & apt genes   | 0.54    | 3874972  | (Z99709) similar to Elongation factor Tu family (contains ATP/GTP binding P-loop); cDNA EST EMBL:D76223 comes from this gene; cDNA EST yk478c5.5 comes from this gene [Caenorhabditis elegans] | 7e-06   |
| 454    | Y08265                                | H.sapiens mRNA for DAN26 protein, partial   | 0.54    | 3875131  | (Z70750) similar to vandate resistance protein transmembranous domains [Caenorhabditis elegans]  | 5e-12   |
| 455    | U89613                                | Hydromantes platycephalus cytochrome b (cytb) gene, mitochondrial gene encoding mitocontrial protein, partial cds | 0.53    | <NONE>   | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 456                                   | AF034597       | Habrobracon hebetor cytochrome oxidase II gene, partial cds; and tRNA-Asp, tRNA His, and tRNA-Lys genes, complete sequence, mitochondrial genes for mitochondrial products  | 0.53    | <NONE>  | <NONE>   | <NONE>  |
| 457                                   | K02653         | Yeast (S.cerevisiae) tau repetitive element and Cys-tRNA.   | 0.53    | <NONE>  | <NONE>   | <NONE>  |
| 458                                   | X53416         | Human mRNA for actin-binding protein (filamin)  | 0.53    | 2134839   | bullous pemphigoid antigen 2 - human   | 6.2     |
| 459                                   | M55545         | Drosophila subobscura alcohol dehydrogenase (Adh) gene, and alcohol dehydrogenase (Adh-dup) gene, complete cds's.   | 0.53    | 2136865   | hair keratin cysteine rich protein - sheep   | 2.1     |
| 460                                   | U19362         | Methanobacterium thermoautotrophicum methylene-tetrahydromethanopterin dehydrogenase (mtd), imidazoleglycerol-phosphate dehydrogenase (hisB), and putative ferredoxin (fdxA) genes, complete cds, orf9 gene, partial cds, orfs... | 0.53    | 731969  | HYPOTHETICAL 91.6 KD PROTEIN IN HXT8-CRT1 INTERGENIC REGION<br>>gi 1078261 pir  S50773 probable membrane protein YJL212c = yeast (Saccharomyces cerevisiae) >gi 496950 (Z34098) ORF [Saccharomyces cerevisiae]<br>>gi 1015596 (Z49487) ORF YJL212c | 0.54    |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 461                                   | AB011527       | Rattus norvegicus mRNA for MEGF1, complete cds                                       | 0.53    | 417037  | GERM CELL-LESS PROTEIN fruit fly (Drosophila melanogaster) >gi 157490 (M97933) germ cell-like protein [Drosophila melanogaster]                                      | 3e-06   |
| 462                                   | U64313         | Bacillus firmus MsyB gene, 5' upstream region and partial cds                        | 0.52    | <NONE>  | <NONE>   | <NONE>  |
| 463                                   | AF008590       | Caenorhabditis elegans paraquat responsive protein (CePqM132) mRNA, complete cds     | 0.52    | <NONE>  | <NONE>   | <NONE>  |
| 464                                   | L10245         | Mus saxicola spermidine/ spermine N1-acetyltransferase (SSAT) gene, complete cds.    | 0.52    | <NONE>  | <NONE>   | <NONE>  |
| 465                                   | AF027173       | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 0.52    | 124263  | INSULIN-LIKE GROWTH FACTOR IB PRECURSOR (IGF-IB) (SOMATOMEDIN C) >gi 69361 pir  IGHU1 B insulin-like growth factor IB precursor - human prepropeptide [Homo sapiens] | 7.7     |
| 466                                   | AL021066       | Caenorhabditis elegans cosmid H31B20, complete sequence [Caenorhabditis elegans]     | 0.52    | 2589162   | (D88451) aldehyde oxidase [Zea mays]   | 6.0     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 467                                   | AF038588       | Porphyra linearis 18S ribosomal RNA gene, 3' partial sequence                         | 0.52    | 1055055   | (U39850) coded for by C. elegans cDNA yk37g1.5; coded for by C. elegans cDNA yk5c9.5; coded for by C. elegans cDNA yk1a9.5; alternatively spliced form of F52C9.8b | 4.6     |
| 468                                   | AE001125       | Borrelia burgdorferi (section 11 of 70) of the complete genome                        | 0.52    | 4115827   | (AB021287) polypeptide [Hepatitis G virus]   | 2.0     |
| 469                                   | AF006640       | Drosophila melanogaster Ste20-like protein kinase mRNA, complete cds                  | 0.52    | 1109830   | (U41534) coded for by C. elegans cDNA CEESI42F; Similar to helicases of SNF2/RAD54 family. [Caenorhabditis elegans]  | 0.002   |
| 470                                   | U90177         | Aplysia californica ubiquitin carboxyl-terminal hydrolase (Ap-uch) mRNA, complete cds | 0.51    | <NONE>  | <NONE>   | <NONE>  |
| 471                                   | Z28304         | S.cerevisiae chromosome XI reading frame ORF YKR079c                                  | 0.51    | <NONE>  | <NONE>   | <NONE>  |
| 472                                   | Z92837         | Caenorhabditis elegans cosmid R03E1, complete sequence [Caenorhabditis elegans]       | 0.51    | 123506  | HYDROPHOBIC SEED PROTEIN (HSP)   | 7.6     |
| 473                                   | D13803         | Mouse mRNA for RecA-like protein MmRad51, complete cds                                | 0.51    | 3327228   | (AB014607) KIA0707 protein [Homo sapiens]  | 4.5     |
| 474                                   | X07187         | Pea hsp21 mRNA  | 0.51    | 3328678   | (AE001299) hypothetical protein [Chlamydia trachomatis]  | 4.4     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 475                                   | S63168         | CCAAT/enhancer-binding protein delta=transcription factor CRP3 homolog [human, prostate carcinoma cell line LNCaP, Genomic 1594 nt] | 0.51    | 1653215   | (D90911) apolipoprotein N-acyltransferase [Synechocystis sp.]                     | 1.2     |
| 476                                   | U67078         | Xenopus laevis C2-HC type zinc finger protein X-MyT1 mRNA, complete cds   | 0.51    | 3850320   | (AF067520) PITSLRE protein kinase beta SV2 isoform [Homo sapiens]                 | 0.17    |
| 477                                   | L38933         | Homo sapiens GT198 mRNA, complete ORF   | 0.51    | 3219965   | HYPOTHETICAL 100.6 KD TRP-ASP REPEATS CONTAINING PROTEIN C2C6.04C IN CHROMOSOME I | 0.059   |
| 478                                   | AF001000       | Lycopersicon esculentum polygalacturonase 1   | 0.50    | <NONE>  | <NONE>  | <NONE>  |
| 479                                   | Z28304         | S.cerevisiae chromosome XI reading frame ORF YKR079c  | 0.50    | <NONE>  | <NONE>  | <NONE>  |
| 480                                   | X97225         | Oncorhynchus keta IGF-II gene   | 0.50    | <NONE>  | <NONE>  | <NONE>  |
| 481                                   | AJ001388       | Homo sapiens, RP58 cDNA for complete mRNA   | 0.50    | <NONE>  | <NONE>  | <NONE>  |
| 481                                   | AJ001388       | Homo Sapiens, RP58 cDNA for complete mRNA   | 0.50    | <NONE>  | <NONE>  | <NONE>  |
| 482                                   | M86626         | P.occultum 23S ribosomal RNA, partial cds.  | 0.50    | <NONE>  | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 483    | U76523                                | Sambucus nigra lectin precursor mRNA, complete cds                                      | 0.50    | 1722856  | CHROMOSOME ASSEMBLY PROTEIN XCAP-E African clawed frog >gi 563814 (U13674) XCAP-e [Xenopus laevis]                                 | 3.2     |
| 484    | AF031663                              | Mus musculus striatin mRNA, complete cds  | 0.50    | 179521   | (M63730) BPAG2 [Homo sapiens]  | 3.2     |
| 485    | U32729                                | Haemophilus influenzae Rd section 44 of 163 of the complete genome                      | 0.50    | 3875699  | (Z92829) F10A3.15 [Caenorhabditis elegans]   | 0.65    |
| 486    | AF067198                              | Dictyostelium discoideum clone 9.10 Tdd-3 and RED repetitive elements, partial sequence | 0.50    | 2494740  | HYPOTHETICAL 28.3 KD PROTEIN IN GBD 5'REGION (ORF4) >gi 2120954 pir  I3956 2 ORF4 - Alcaligenes eutrophus >gi 695274 (L36817) ORF4 | 0.008   |
| 487    | M23442                                | Human interleukin 4 (IL-4) gene, complete cds.  | 0.49    | <NONE>   | <NONE>   | <NONE>  |
| 488    | U16367                                | Caenorhabditis elegans POU homeobox protein CEH-18 (ceh-18) mRNA, complete cds.         | 0.47    | 3786409  | (AF098499) contains similarity to Saccharomyces cerevisiae MAF1 protein (GB:U19492) [Caenorhabditis elegans]                       | 8.9     |
| 489    | AF001000                              | Lycopersicon esculentum polygalacturonase 1   | 0.45    | <NONE>   | <NONE>   | <NONE>  |
| 490    | Z18920                                | Yersinia enterocolitica wbb gene cluster  | 0.41    | <NONE>   | <NONE>   | <NONE>  |
| 491    | D86983                                | Human mRNA for KIAA0230 gene, partial cds   | 0.35    | 206712   | (M64793) salivary proline-rich protein [Rattus norvegicus]   | 4e-05   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 492    | AF064030                              | Helianthus tuberosus lectin 2 mRNA, complete cds   | 0.33    | <NONE>   | <NONE>  | <NONE>  |
| 493    | AF067083                              | Vitreoscilla sp. outer membrane protein homolog gene, complete cds; Trp repressor binding protein gene, partial cds; and unknown genes | 0.33    | 401553   | HYPOTHETICAL 24.5 KD PROTEIN IN NADB-SRMB INTERGENIC REGION   | 8.3     |
| 494    | Y15520                                | Papio hamadryas anubis gene encoding fertilin alpha-II   | 0.29    | 2408049  | (Z99164) hypothetical protein   | 3.1     |
| 495    | U33475                                | Alestes sp. ependymin mRNA, partial cds  | 0.28    | 3913078  | ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR HOMOLOG (DARNT) (TANGO PROTEIN) transcription factor [Drosophila melanogaster] | 1.4     |
| 496    | D67603                                | Mouse DNA for 8-oxodGTPase, complete cds   | 0.22    | <NONE>   | <NONE>  | <NONE>  |
| 497    | U82386                                | Methanococcus jannaschii section 145 of 150 of the complete genome   | 0.22    | 2209261  | (U51222) p40 [Streptomyces halstedii]   | 8.3     |
| 498    | U82386                                | Malurus cyaneus microsatellite McyU2   | 0.22    | 992631   | (U29131) Mg-chelataase subunit [Synechocystis sp.]  | 0.56    |
| 499    | Z49625                                | S.cerevisiae chromosome X reading frame ORF YJR125c  | 0.21    | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 500    | U64830                                | Dictyostelium discoideum AX2 protein tyrosine kinase gene, complete cds. | 0.21    | <NONE>   | <NONE>   | <NONE>  |
| 501    | M24543                                | Human prostate-specific antigen (PA) gene, complete cds.                 | 0.21    | 2764859  | (X97918) gene 12.1 [Bacteriophage SPP1]  | 6.0     |
| 502    | X87618                                | B.taurus mRNA for thrombospondin (partial) 2162 bp                       | 0.21    | 2146000  | u0002b protein - Mycobacterium tuberculosis tuberculosis] >gi 1694863 gnl PID e 283373 (Z83018) hypothetical protein Rv2968c [Mycobacterium tuberculosis]  | 3.5     |
| 503    | X71591                                | B.taurus microsatellite sequence INRA048                                 | 0.21    | 1354453  | (U52830) orf [Homo sapiens]  | 2.7     |
| 504    | X57808                                | Human germline immunoglobulin lambda light chain gene                    | 0.21    | 2119158  | procollagen type V alpha 2 - mouse >gi 309181  | 2.7     |
| 505    | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds               | 0.21    | 2497139  | HYPOTHETICAL 78.8 KD PROTEIN IN ABF2-CHL12 INTERGENIC REGION >gi 1078003 pir  S528 35 hypothetical protein YMR075w - yeast (Saccharomyces cerevisiae) >gi 763022 (Z48952) unknown [Saccharomyces cerevisiae] | 2.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 506    | U84216                                | Mycobacterium fortuitum plasmid pJAZ38 replication protein Rep (rep) gene, complete cds | 0.21    | 2499087  | UDP-GLUCOSE: GLYCOPROTEIN GLUCOSYLTRANSF ERASE PRECURSOR (DUGT) glucosyltransferase - fruit fly (Drosophila sp.) glucosyltransferase precursor [Drosophila melanogaster] | 0.003   |
| 507    | U31463                                | Rattus norvegicus nonmuscle myosin heavy chain-A mRNA, complete cds.                    | 0.21    | 3880111  | (Z81130) predicted using Genefinder  | 0.002   |
| 508    | X51508                                | Rabbit mRNA for aminopeptidase N (partial)  | 0.21    | 630864   | LRR47 protein - fruit fly (Drosophila melanogaster) >gi 415947 (X75760) LRR47 [Drosophila melanogaster]  | 1e-06   |
| 509    | AF086476                              | Homo sapiens full length insert cDNA clone ZD88F12                                      | 0.20    | <NONE>   | <NONE>   | <NONE>  |
| 510    | AF077006                              | Helicobacter pylori plasmid pHPM186, complete sequence                                  | 0.20    | <NONE>   | <NONE>   | <NONE>  |
| 511    | X75480                                | E.gunnii CAD gene.  | 0.20    | <NONE>   | <NONE>   | <NONE>  |
| 512    | X75036                                | T.raestivum mitochondrial nad7 gene for NADH dehydrogenase subunit 7                    | 0.20    | <NONE>   | <NONE>   | <NONE>  |
| 513    | D90875                                | E.coli genomic DNA, Kohara clone #422(55.5-55.8 min.)                                   | 0.20    | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION   | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 514    | Z68343                                | Caenorhabditis elegans cosmid F59B8, complete sequence [Caenorhabditis elegans] | 0.20    | <NONE>   | <NONE>  | <NONE>  |
| 515    | X62486                                | M.musculus V alpha 11.1 gene 5'-region  | 0.20    | <NONE>   | <NONE>  | <NONE>  |
| 516    | AF040651                              | Caenorhabditis elegans cosmid W04H10  | 0.20    | 1170683  | PHOSPHORYLASE B KINASE ALPHA REGULATORY CHAIN, SKELETAL MUSCLE ISOFORM (PHOSPHORYLASE KINASE ALPHA M SUBUNIT) >gi 2135923 pir  I38111 phosphorylase kinase (EC 2.7.1.38) - human >gi 791043 | 7.4     |
| 517    | U10470                                | Pseudomonas fluorescens PHA depolymerase (phaZ) gene, complete cds.             | 0.20    | 3721862  | (AB016024) Pfj2 [Plasmodium falciparum]   | 1.9     |
| 518    | D83778                                | Human mRNA for KIAA0194 gene, partial cds                                       | 0.20    | 126363   | LAMININ ALPHA-1 CHAIN PRECURSOR precursor - human   | 0.65    |
| 519    | S43579                                | c-scr=pp60c-src, sdr=src downstream region                                      | 0.20    | 4159887  | (AC004908) similar to ribosomal protein L23a; similar to P29316 (PID:g132848) [Homo sapiens]  | 0.52    |
| 520    | U07357                                | Mus musculus Balb/c brain-specific kinase (Bsk) mRNA, complete cds.             | 0.20    | 206712   | (M64793) salivary proline-rich protein [Rattus norvegicus]  | 0.51    |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 521    | AF034460                              | Penicillium thomii internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence. | 0.20    | 114136   | AMINO-ACID ACETYLTRANSFERASE Pseudomonas aeruginosa >gi 151036 (M38358) N-acetylglutamate synthase [Pseudomonas aeruginosa]   | 0.39    |
| 522    | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds  | 0.20    | 2842674  | POU DOMAIN CLASS 2, ASSOCIATING FACTOR 1 (B-CELL-SPECIFIC COACTIVATOR OBF-1) (OCT BINDING FACTOR 1) (BOB-1) (OCA-B) Bob1, B-cell-specific - mouse >gi 1881818 bbs 1798 52 mBob1=B-cell specific transcriptional coactivator line J558L, Peptide, 256 aa] >gi 1353792 (U43788) Oct binding factor 1 [Mus musculus] | 0.073   |
| 523    | X95971                                | S.lividans groEL2 gene  | 0.20    | 3925277  | (AL032643) similar to Uncharacterized protein family UPF0034, Double-stranded RNA binding motif; cDNA EST yk489b3.5 comes from this gene; cDNA EST yk439g7.5 comes from this gene [Caenorhabditis elegans]  | 4e-19   |
| 524    | L41502                                | Ovis aries vasopressin V1 receptor (V1R) gene, complete cds   | 0.19    | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|---|---------|--|-------------|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION | P VALUE |
| 525    | J03885                                | K.pneumoniae oxalacetate decarboxylase alpha subunit gene, complete cds.                          | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 526    | AE001451                              | Helicobacter pylori, strain J99 section 12 of 132 of the complete genome                          | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 527    | D88084                                | Pedicularis verticillata chloroplast DNA, intergenic region between trnT(UGU) and trnL(UAA)5'exon | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 528    | U67599                                | Methanococcus jannaschii section 141 of 150 of the complete genome                                | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 529    | J05500                                | Human beta-spectrin (SPTB) mRNA, complete cds.  | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 530    | Y10137                                | M.mycoides ftsY gene homologue and gene encoding hypothetical protein                             | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 531    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds              | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 532    | D43805                                | Mouse thymic stromal cell mRNA for TLSF-beta, complete cds  | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 533    | AJ012585                              | Tetrahymena thermophila macronuclear gene encoding ribosomal protein L3, exons 1-2                | 0.19    | <NONE>   | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 534    | X51475                                | Brassica napus 5-enolpyruvylshikimate-3-phosphate synthase gene                  | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 535    | AF074386                              | Sambucus nigra hevin-like protein mRNA, complete cds                             | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 536    | Z49625                                | S.cerevisiae chromosome X reading frame ORF YJR125c                              | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 537    | X63741                                | H.sapiens pilot mRNA   | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 538    | Y11255                                | O.latipes mRNA for annexin max4  | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 539    | L63537                                | Oncorhynchus mykiss (clone Jb-10) beta-2 microglobulin (B2m) mRNA, complete cds. | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 540    | X70903                                | N.tobacum T92 gene for auxin-binding protein                                     | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 541    | U61958                                | Caenorhabditis elegans cosmid C25A8  | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 542    | U33959                                | Macaca fascicularis fertilin beta mRNA, complete cds                             | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 543    | Z49835                                | H.sapiens mRNA for protein disulfide isomerase                                   | 0.19    | 2113940  | (Z95556) hypothetical protein Rv2507  | 9.4     |
| 544    | AF035458                              | Spinacia oleracea heat shock 70 protein, complete cds                            | 0.19    | 267293   | PROBABLE E4 PROTEIN papillomavirus (type 1) >gi 61015 (X62844) E4 gene product [Pygmy chimpanzee papillomavirus type 1] | 9.4     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 545                                   | U23441         | Tetrahymena thermophila B internal deletion sequence.  | 0.19    | 3877185   | (Z66563) F46C3.2 [Caenorhabditis elegans]   | 9.3     |
| 546                                   | U53921         | Pneumocystis carinii major surface glycoprotein  | 0.19    | 3548901   | (AF052502) DA26 homolog [Epiphyas postvittana nucleopolyhedrovirus]   | 9.3     |
| 547                                   | L11002         | Rat ankyrin binding glycoprotein-1 related mRNA sequence.  | 0.19    | 3337352   | (AC004481) putative chromatin structural protein Supt5hp  | 9.1     |
| 548                                   | U67560         | Methanococcus jannaschii section 102 of 150 of the complete genome   | 0.19    | 3183689   | (Y13585) serotonin receptor 4 [Cavia porcellus]   | 8.7     |
| 549                                   | U18424         | Mus musculus bacteria binding macrophage receptor MARCO mRNA, complete cds.                                | 0.19    | 3659853   | (AF089083) complement component C1qB like protein   | 7.1     |
| 550                                   | X66467         | C.albicans sec18 gene  | 0.19    | 1326385   | (U58751) C07G1.7 gene product [Caenorhabditis elegans]  | 6.9     |
| 551                                   | AF003487       | Syngaster lepidus 16S ribosomal RNA gene, partial sequence   | 0.19    | 3122039   | DIHYDROPYRIMIDINASE (DHPASE) dihydropyrimidinase - rat<br>>gi 1378019 gnl PID d1010479                          | 6.9     |
| 552                                   | J05087         | Rat calmodulin-sensitive plasma membrane Ca <sup>2+</sup> -transporting ATPase (PMCA3) mRNA, complete cds. | 0.19    | 422462  | hypothetical protein - fruit fly (Drosophila melanogaster)<br>>gi 296434 (X68408) ORF [Drosophila melanogaster] | 5.3     |
| 553                                   | AF080464       | Homo sapiens glutamate oxaloacetate transaminase   | 0.19    | 3024834   | PROBABLE E4 PROTEIN >gi 790898 position 3286..3288 is first state condon; putative                              | 5.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 554    | U78876                                | Human MEK kinase 3 mRNA, complete cds                                  | 0.19    | 1710445  | (U78083) unknown [Emericella nidulans]  | 5.3     |
| 555    | AB009077                              | Vigna radiata mRNA for proton pyrophosphatase, complete cds            | 0.19    | 3256922  | (AP000002) 256aa long hypothetical protein [Pyrococcus horikoshii]  | 5.1     |
| 556    | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds             | 0.19    | 4226159  | (AF125463) contains similarity to BTB (also known as BR-C/Ttk) domains (Pfam:PF00651, Score=62.8, E=7.6e-15, N=1) [Caenorhabditis elegans]  | 4.1     |
| 557    | AE000392                              | Escherichia coli K-12 MG1655 section 282 of 400 of the complete genome | 0.19    | 3645960  | (AL031583) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=47.46; 1-evidence_end; 2-evidence=predicted by match; 2-match_accession=SWISS-PROT:P23792; 2-match_description=DISCONNECTED PROTEIN.; 2-matc... | 4.0     |
| 558    | AE000392                              | Escherichia coli K-12 MG1655 section 282 of 400 of the complete genome | 0.19    | 3645960  | (AL031583) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=47.46; 1-evidence_end; 2-evidence=predicted by match; 2-match_accession=SWISS-PROT:P23793; 2-match_description=DISCONNECTED PROTEIN.; 2 matc... | 4.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 559    | L81774                                | Homo sapiens (subclone 3_d1 from P1 H25) DNA sequence                                   | 0.19    | 4001725  | (AB015981) MnhA [Staphylococcus aureus]   | 3.0     |
| 560    | AL021108                              | Drosophila melanogaster cosmid clone 137E7  | 0.19    | 4001688  | (AB015718) protein kinase [Homo sapiens]  | 3.0     |
| 561    | AB001510                              | Carabus leptoplesioides mitochondrial DNA for NADH dehydrogenase subunit 5, partial cds | 0.19    | 3758855  | (Z98551) MAL3P6.11 [Plasmodium falciparum]  | 2.4     |
| 562    | AF069696                              | Egernia stokesii clone EST1 microsatellite  | 0.19    | 3328994  | (AE001326) Amino Acid (Branched) Transport [Chlamydia trachomatis]  | 2.4     |
| 563    | X64144                                | F.pringlei ppcA1 gene for phosphoenolpyruvate carboxylase                               | 0.19    | 3242974  | (AF069555) G protein-coupled receptor p2y3 [Meleagris gallopavo]  | 2.3     |
| 564    | U56897                                | Human immunodeficiency virus type 1 gag polyprotein (gag) gene, partial cds             | 0.19    | 2257710  | (U73041) resolvase-like protein [Thiobacillus ferrooxidans]   | 2.3     |
| 565    | U57975                                | Danio rerio Notch homologue 3 mRNA, complete cds  | 0.19    | 3874971  | (Z99/09) similar to NAD dependent epimerase/dehydratase family; cDNA EST EMBL:C10103 comes from this gene; cDNA EST EMBL:D66400 comes from this gene; cDNA EST EMBL:D70143 comes from this gene; cDNA EST yk493h11.3 comes from ... | 1.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 566    | Y12502                                | R.norvegicus mRNA for factor XIIIa   | 0.19    | 2133693  | masquerade precursor - fruit fly (Drosophila melanogaster) >gi 665545 (U18130) masquerade [Drosophila melanogaster] >gi 1095942 prf 2110286A masquerade gene | 1.8     |
| 567    | S82470                                | BB1=malignant cell expression-enhanced gene/ tumor progression-enhanced gene [human, UM-UC-9 bladder carcinoma cell line, mRNA, 1897 nt]   | 0.19    | 2444026  | (U77783) N-methyl-D-aspartate receptor 2D subunit precursor [Homo sapiens]   | 1.8     |
| 568    | U97408                                | Caenorhabditis elegans cosmid F48A9  | 0.19    | 542433   | 225K protein - Babesia bovis (fragment)  | 1.8     |
| 569    | U10470                                | Pseudomonas fluorescens PHA depolymerase (phaZ) gene, complete cds.  | 0.19    | 3721862  | (AB016024) Pfj2 [Plasmodium falciparum]  | 1.7     |
| 570    | M88160                                | Ovis aries MAF214 locus polymorphic dinucleotide repeat.   | 0.19    | 1293816  | (U56963) T13A10.5 gene product [Caenorhabditis elegans]  | 1.4     |
| 571    | Aj131336                              | Horcus ranatus mRNA for pollen allergen (Hol i 2, group II) >:: emb AJ131339 LIT131339 Lolium italicum mRNA for pollen allergen (Lol i 2, group II) > allergen (Poa p 2, group II) >:: emb AJ131338 TAE131338 Triticum aestivum mRNA for pollen allergen (Tri a 2, group II) | 0.19    | 3880447  | (AL032675) predicted using Genefinder  | 0.82    |

| Nearest Neighbor (BlastN vs. Genbank) |             |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-------------|--|---------|--|--|---------|
| SEQ ID                                | ACCES- SION | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 572                                   | X84036      | S.cerevisiae ARG8 and CDC33 genes  | 0.19    | 3882041  | (AJ010405) hypothetical protein  | 0.62    |
| 573                                   | U57058      | Human WD protein IR10 pre-mRNA, partial cds  | 0.19    | 631302   | mucin - human >gi 501033 (U14383) mucin [Homo sapiens]   | 0.60    |
| 574                                   | AF034460    | Penicillium thomii internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence | 0.19    | 114136   | AMINO-ACID ACETYLTRANSFERASE Pseudomonas aeruginosa >gi 151036 (M38358) N-acetylglutamate synthase [Pseudomonas aeruginosa]                            | 0.35    |
| 575                                   | U95098      | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds   | 0.19    | 105270   | alpha-2-adrenergic receptor - human name 'ADRA2R' [Homo sapiens]   | 0.27    |
| 576                                   | AG001475    | Homo sapiens genomic DNA, 21q region, clone: 125H6N2   | 0.19    | 94977  | hypothetical protein 3 - Pseudomonas sp. (DSM 6898(plasmid pKB740 >gi 45867 (X66604) ORF3  | 0.16    |
| 577                                   | M63284      | Mouse IgG receptor (beta-Fc-gamma-RII) gene, exons 9 and 10, clones lambda-Fc(3.2,93).   | 0.19    | 3024681  | TRANSCRIPTION INITIATION FACTOR TFIID 135 KD SUBUNIT (TAFII-135) (TAFII135) (TAFII-130) of RNA polymerase II transcription factor TFIID [Homo sapiens] | 0.088   |
| 578                                   | U38241      | Pseudomonas aeruginosa orotate phosphoribosyl transferase (pyrE), catabolite repression control protein (crc) and RNasePH (rph) genes, complete cds                          | 0.19    | 3044086  | (AF055904) unknown [Myxococcus xanthus]  | 0.052   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 579    | AF039734                              | Lontra longicaudis transthyretin intron 1, partial sequence                | 0.19    | 322759   | pistil extensin-like protein (clone pMG14) - common tobacco (fragment) >gi 19927 (Z14015) pistil extensin like protein [Nicotiana tabacum]  | 0.030   |
| 580    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                    | 0.19    | 2147194  | collagen - Paralvinella grasslei  | 0.002   |
| 581    | AB004232                              | Drosophila melanogaster mRNA for DAD polypeptide, complete cds             | 0.19    | 2498765  | PEROXISOMAL MEMBRANE PROTEIN PEX16 [lipolytica]   | 0.002   |
| 582    | AF098919                              | Gallus gallus alpha-globin gene domain 5' region                           | 0.19    | 1086863  | (U41272) T03G11.6 gene product [Caenorhabditis elegans]   | 4e-05   |
| 583    | AE001457                              | Helicobacter pylori, strain J99 section 18 of 132 of the complete genome   | 0.19    | 2924552  | (AL022018) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=165.48; 1-evidence_end; 2-evidence=predicted by match; 2-match_accession=AA264666; 2-match_description=LD08351.5prime LD Drosophila melanoga... | 3e-05   |
| 584    | L10329                                | Plasmid RP4 traE gene, 3' end; traD gene, complete cds; traF gene, 5' end. | 0.19    | 3878117  | (Z49068) mitochondrial carrier protein  | 8e-07   |
| 585    | AE001155                              | Borrelia burgdorferi (section 41 of 70) of the complete genome             | 0.19    | 861276   | (U28739) similar to TPR domains in e.g. yeast STI1 protein [Caenorhabditis elegans]   | 2e-12   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 586    | U49979                                | Orf virus E10R homolog gene, partial cds, and DNA polymerase gene, complete cds                 | 0.19    | 3850072  | (AL033385) dna-directed rna polymerase iii subunit [Schizosaccharomyces pombe]   | 1e-15   |
| 587    | U88155                                | Xenopus laevis RanGTPase activating protein   | 0.19    | 995714   | (X91258) pid:e198503 [Saccharomyces cerevisiae]  | 4e-16   |
| 588    | AF061854                              | Schizosaccharomyces pombe Clr4p (clr4) gene, complete cds                                       | 0.19    | 3242750  | (AC005164) match to ESTs AA731149 (NID:g2140138), AA731908 (NID:g2752719), AA287837 (NID:g1933519), AA262811 (NID:g1898382), and AA825820 (NID:g2899132) | 5e-19   |
| 589    | M23865                                | S.cerevisiae CHS2 gene encoding chitin synthase.  | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 590    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds   | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 591    | AF067610                              | Caenorhabditis elegans cosmid F41A4   | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 592    | AF036329                              | Homo sapiens gonadotropin-releasing hormone precursor, second form (GnRH-II) gene, complete cds | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 593    | Z49216                                | H.sapiens mitoxantrone-resistance associated mRNA   | 0.18    | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 594    | X02167                                | Torulopsis glabrata mitochondrial DNA for tRNA-Thr,-His and -Glu upstream of cytochrome b gene                                | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 595    | Z31561                                | R.communis (Carmencita) Scr1 mRNA for sucrose carrier   | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 596    | L81692                                | Homo sapiens (subclone 2_c9 from P1 H56) DNA sequence   | 0.18    | 1346575  | 55 KD ERYTHROCYTE MEMBRANE PROTEIN   | 8.4     |
| 597    | X57310                                | Nocardia lactamdurans pcbAB and pcbC genes for alpha-aminoadipyl-L-cysteinyl-D-valine synthetase and isopenicillin N synthase | 0.18    | 126404   | SEED LIPOXYGENASE-2 (L-2) soybean >gi 170014 (J03211) lipoxygenase (EC 1.13.11.12)   | 6.5     |
| 598    | U18315                                | Sus scrofa parathyroid receptor (PTH) mRNA, complete cds  | 0.18    | 1022323  | (X04647) collagen alpha-2(IV) chain [Mus musculus]   | 3.8     |
| 599    | AL010158                              | Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 3-85, complete sequence                                  | 0.18    | 2506816  | VERSICAN CORE PROTEIN PRECURSOR PROTEOGLYCAN COR PROTEIN 2) (GLIAL HYALURONATE-BINDING PROTEIN) (GHAP) >gi 608515 (U16306) chondroitin sulfate proteoglycan versican V0 splice-variant precursor peptide | 3.7     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 600                                   | AB005287       | Bos taurus mRNA for thrombospondin 1, complete cds         | 0.18    | 2146000   | U0002b protein - Mycobacterium tuberculosis tuberculosis] >gi 1694863 gnl PID e 283373 (Z83018) hypothetical protein Rv2968c [Mycobacterium tuberculosis]                            | 2.9     |
| 601                                   | AL021108       | Drosophila melanogaster cosmid clone 137E7                 | 0.18    | 3483032   | (AL031371) hypothetical protein SC4G.06 [Streptomyces coelicolor]  | 2.9     |
| 602                                   | U57975         | Danio rerio Notch homologue 3 mRNA, complete cds           | 0.18    | 85719   | Collagen alpha 1'(II) chain precursor - African clawed frog  | 1.7     |
| 603                                   | M30124         | P.aeruginosa autonomously replicating sequence.            | 0.18    | 3878017   | (AL021387) similar to Zinc finger, C4 type (two domains); cDNA EST yk452f4.5 comes from this gene; cDNA EST EMBL:T00774 comes from this gene receptor NHR-3 (Caenorhabditis elegans] | 1.3     |
| 604                                   | X54965         | G.sp alpha 5HR DNA   | 0.18    | 134304  | STEM CELL PROTEIN chicken >gi 62845 (X63371) transforming capacity [Gallus gallus]   | 1.3     |
| 605                                   | U95098         | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds | 0.18    | 1628403   | (X98893) hTAFII68 [Homo sapiens] splicing [Homo sapiens]   | 1.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 606    | U20793                                | Oryctolagus cuniculus renal sodium-dependent phosphate transporter type II mRNA, complete cds. | 0.18    | 1705984  | 92 KD TYPE IV COLLAGENASE PRECURSOR IV, 92K, precursor - rat >gi 1022784 (U36476) 92-kDa type IV collagenase [Rattus norvegicus]   | 1.2     |
| 607    | U23427                                | Human cholecystokinin type A receptor (CCK-A) gene, exons 1 and 2.                             | 0.18    | 3261734  | (Z94752) hypothetical protein Rv1004c  | 0.97    |
| 608    | U49953                                | Rattus norvegicus protein kinase MUK2 mRNA, complete cds                                       | 0.18    | 551238   | (X81847) pectate lyase 1 [Erwinia carotovora]  | 0.43    |
| 609    | J00182                                | Human alpha globin gene cluster on chromosome 16: zeta gene.                                   | 0.18    | 1585259  | traJ gene [Amycolatopsis methanolica]  | 0.41    |
| 610    | X62513                                | M.gallopavo gene for metallothionein   | 0.18    | 2494740  | HYPOTHETICAL 28.3 KD PROTEIN IN GBD 5'REGION (ORF4) >gi 2120954 pir  I3956 2 ORF4 - Alcaligenes eutrophus >gi 695274 (L36817) ORF4 | 0.31    |
| 611    | X04862                                | Goat embryonic alpha globin gene zeta exons 2-3  | 0.18    | 86837  | androgene receptor B - human   | 0.082   |
| 612    | M12450                                | Rat vitamin D binding protein (DBP) mRNA, complete cds   | 0.18    | 4210432  | (AJ130783) APC2 protein [Mus musculus]   | 0.038   |
| 613    | AF038539                              | Mus musculus muscle NSP-like 1 (Nsp11) mRNA, complete cds                                      | 0.18    | 3297877  | (AJ224868) GNAS1 (Homo sapiens)  | 0.029   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 614    | X78401                                | Bacteriophage P22 right operon, orf 48, replication genes 18 and 12, nin region genes, ninG phosphatase, late control gene 23, orf 60, complete cds, late control region, start of lysis gene 13 | 0.18    | 1123087  | (U42436) C49H3.3 gene product [Caenorhabditis elegans]  | 0.009   |
| 615    | D38754                                | Pig mRNA for inter-alpha-trypsin inhibitor heavy-chain H1, complete cds  | 0.18    | 1397275  | (U61947) C06G3.8 gene product [Caenorhabditis elegans]  | 7e-06   |
| 616    | X51508                                | Rabbit mRNA for aminopeptidase N (partial)   | 0.18    | 630864   | LRR47 protein - fruit fly (Drosophila melanogaster) >gi 415947 (X75760) LRR47 [Drosophila melanogaster]   | 6e-07   |
| 617    | X54850                                | S.kluyveri linear plasmid pSKL DNA for open reading frams 1-1-   | 0.18    | 3183405  | HYPOTHETICAL 11.3 KD PROTEIN C2C6.07 IN CHROMOSOME I >gi 2370504 gnl PID e 339194 pombe] >gi 3451305 gnl PID e 1316730 (AL031324) very hypothetical protein [Schizosaccharomyces pombe] | 2e-08   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 618    | L21954                                | Human peripheral benzodiazepine receptor gene, exon 4.  | 0.18    | 3925211  | (AL03626) cDNA EST<br>EMBL:D70654 comes from this gene; cDNA EST<br>EMBL:Z14359 comes from this gene; cDNA EST<br>EMBL:D33409 comes from this gene; cDNA EST<br>EMBL:D36239 comes from this gene; cDNA EST<br>EMBL:Z14766 comes from this gene... | 4e-09   |
| 619    | U09355                                | Oryctolagus cuniculus protein phosphatase 2A1 B gamma subunit (skeletal muscle isolate) mRNA, complete cds. | 0.18    | 3947877  | (AL034382) putative mitosis and maintenance of ploidy protein (Schizosaccharomyces pombe)   | 8e-11   |
| 620    | X58715                                | T.cruzi hsp70 mRNA for 70 kDa heat shock protein, partial cds   | 0.18    | 3024081  | MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES (MLCK) (CONTAINS: TELOKIN)   | 9e-12   |
| 621    | AF060195                              | Mus musculus proteasome regulator PA28 beta subunit gene, complete cds                                      | 0.18    | 861276   | (U28739) similar to TPR domains in e.g. yeast STI1 protein [Caenorhabditis elegans]   | 1e-14   |
| 622    | L27235                                | Methylobacterium extorquens serine cycle proteins   | 0.18    | 2688949  | (AF027208) AC133 antigen [Homo sapiens]   | 1e-14   |
| 623    | AF006573                              | Drosophila virilis maltase 1 (Mav1) and maltase 2 (Mav2) genes complete cds                                 | 0.18    | 2500558  | PUTATIVE RIBONUCLEASE III (RNASE III)<br>>gi 38764201 gnl PID e1346063  | 2e-23   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 624                                   | AF001782       | Staphylococcus aureus strain SA502A AgrB   | 0.17    | <NONE>  | <NONE>  | <NONE>  |
| 625                                   | AJ223364       | Homo sapiens germline DNA upstream of Jkappa locus                                   | 0.17    | <NONE>  | <NONE>  | <NONE>  |
| 626                                   | J03059         | Human glucocerebrosidase (GCB) gene, complete cds                                    | 0.17    | <NONE>  | <NONE>  | <NONE>  |
| 627                                   | AB008860       | Fugu rubripes Ca12 gene for pheromone receptor, complete cds                         | 0.17    | 2198849   | (AF004900) E3KARP [Homo sapiens]<br>>gi 2665826 (AF035771) Na+H+ exchanger regulatory factor 2 [Homo sapiens] factor 2 [Homo sapiens]<br>>gi 3618353 gnl PID d1034182 exchanger isoform A3 [Homo sapiens] | 7.8     |
| 628                                   | AF027174       | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 0.17    | 539355  | SCD25 protein (version 1) - yeast   | 7.5     |
| 629                                   | AF059650       | Homo sapiens histone deacetylase 3 (HDAC3) gene, complete cds                        | 0.17    | 482118  | Hypothetical protein C15H7.1 - Caenorhabditis elegans   | 4.5     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 630    | AF059650                              | Homo sapiens histone deacetylase 3 (HDAC3) gene, complete cds           | 0.17    | 465932   | HYPOTHETICAL 83.2 KD PROTEIN F58A4.11 IN CHOMOSOME III >gi 38742887 gnl PID e1344088 EST EMBL:C12577 comes from this gene; cDNA EST yk227e7.5 comes from this gene; cDNA EST yk303d1.5 comes from this gene; cDNA EST yk314c12.5 comes from this gene; cDNA... EMBL:C11886 comes from this gene: cDNA EST EMBL:C12577 comes from this gene; cDNA EST yk22737.5 comes from this gene; cDNA EST yk303d1.5 comes from this gene; cDNA EST yk314c12.5 comes from this gene; cDNA... | 4.4     |
| 631    | X55065                                | Chinese hamster metallothionein II gene                                 | 0.17    | 3687237  | (AC005169) putative Cys3His zinc-finger protein   | 1.5     |
| 632    | U15280                                | Rattus norvegicus oxytocin receptor (OTR) gene, exon 3 and complete cds | 0.17    | 542565   | Cyclin E type II - fruit fly [Drosophila melanogaster] <gi 429168 (X75027) Drosophila cyclin E type II [Drosophila melanogaster]  | 0.45    |
| 633    | X04862                                | Goat embryonic alpha globin gene zeta exons 2-3                         | 0.17    | 86837  | Androgen receptor B - human   | 0.080   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 634                                   | AL010222       | Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 4-09, complete sequence    | 0.17    | 1177322   | (X95466) CPG2 protein [Rattus norvegicus] <gi 1588593 prf 2208498A plasticity-related gene [Rattus norvegicus]   | 7e-07   |
| 635                                   | X60111         | H.sapiens mRNA for MRP-1  | 0.17    | 3237306   | (U92715) breast cancer antiestrogen resistance 3 protein   | 3e-09   |
| 636                                   | U49979         | Orf virus E10R homolog gene, partial cds, and DNA polymerase gene, complete cds                 | 0.17    | 3850072   | (AL033385) dan-directed rna polymerase iii subunit [Schizosaccharomyces pombe]   | 3e-09   |
| 637                                   | AF006573       | Drosophila virilis maltase 1 (Mav1) and maltase 2 (Mav2) genes, complete cds                    | 0.17    | 2500558   | PUTATIVE RIBONUCLEASE III (RNASE III) >gi 3876420 gnl PID e1346063 (Z81070) similar to ribonuclease [Caenorhabditis elegans]                             | 2e-29   |
| 638                                   | AE001141       | Borrelia burgdorferi (section 27 of 70) of the complete genome                                  | 0.16    | 1850327   | (U52370) fertilin beta [Homo sapiens]  | 2.3     |
| 639                                   | M72980         | Anthonomus grandis vitellogenin gene (VTG), complete cds.                                       | 0.12    | 3242750   | (AC005164) match to ESTs AA731149 (NID:g2140138), AA731908 (NID:g2752719), AA287837 (NID:g1933519), AA262811 (NID:g1898382), and AA825820 (NID:g2899132) | 2e-56   |
| 640                                   | AF023532       | Simulium vittatum ATPase 6 gene, mitochondrial gene encoding mitochondrial protein, partial cds | 0.11    | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION   | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 641    | U76523                                | Sambucus nigra lectin precursor mRNA, complete cds  | 0.10    | 3482965  | (AL031369) putativ protein  | 0.49    |
| 642    | AF001596                              | Danio rerio mRNA for opioid receptor homologue  | 0.099   | 1706694  | LANOSTEROL SYNTHASE 5.4.99.7) - fission yeast (Schizosaccharomyces pombe) | 2.3     |
| 643    | U26341                                | Oryctolagus cuniculus Na and Cl dependent betaine transporter mRNA, complete cds.         | 0.099   | 2645804  | (AF033381) betaine homocysteine methyl transferase [Mus musculus]         | 0.59    |
| 644    | M11633                                | Bacteriophage Cp-5 (S.pneumoniae) 3' inverted terminal repeat.                            | 0.082   | 2314695  | (AE000649) type IIS restriction enzyme R and M protein                    | 4.3     |
| 645    | X74103                                | Streptomyces sp. Gene for alkaline serine protease I                                      | 0.073   | 1314734  | (U54641) 220 kDa silk protein [Chironomus thummi]                         | 6.3     |
| 646    | Z72509                                | Caenorhabditis elegans cosmid F32G8, complete sequence [Caenorhabditis elegans]           | 0.072   | <NONE>   | <NONE>  | <NONE>  |
| 647    | X70282                                | X.laevis xanf-1 gene  | 0.070   | 3851202  | (AC005954) ZO-3 [Homo sapiens] [Homo sapiens]                             | 0.40    |
| 648    | Z69906                                | Human DNA sequence from cosmid E141E2, on chromosome 22, complete sequence [Homo sapiens] | 0.069   | <NONE>   | <NONE>  | <NONE>  |
| 649    | AF056940                              | Drosophila virilis retrotransposon Tv1, complete sequence                                 | 0.069   | 2246532  | (U93872) ORF 73, contains large complex repeat CR 73                      | 5e-12   |
| 650    | AJ001151                              | Homo sapiens genomic sequence   | 0.068   | <NONE>   | <NONE>  | <NONE>  |
| 651    | X54455                                | Bacteriophage BF23 gene 17 and gene 18  | 0.067   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 652    | X87936                                | P.pinea internal transcribed spacers 1 & 2 of ribosomal DNA                                   | 0.067   | 2459733  | (U95374) aldehyde dehydrogenase [Haloferax volcanii]           | 4.3     |
| 653    | AF019236                              | Dictyostelium discoideum TipD (tipD) gene, complete cds                                       | 0.067   | 3882275  | (AB018320) KIAA0777 protein [Homo sapiens]                     | 1.1     |
| 654    | X90592                                | O.cuniculus mRNA for p53 protein  | 0.067   | 1703275  | METHIONINE AMINOPEPTIDASE 2 (METAP 2) GLYCOPROTEIN) (P67)      | 0.29    |
| 655    | U41805                                | Mus musculus putative TI/ST2 receptor binding protein precursor mRNA, partial cds             | 0.067   | 642518   | (U17326) neuronal nitric oxide synthase [Homo sapiens]         | 0.29    |
| 656    | AB007881                              | Homo sapiens KIAA0421 mRNA, partial cds   | 0.066   | <NONE>   | <NONE>   | <NONE>  |
| 657    | AL010213                              | Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 3-109, complete sequence | 0.066   | <NONE>   | <NONE>   | <NONE>  |
| 658    | AB014546                              | Homo sapiens mRNA for KIAA0646 protein, complete cds  | 0.066   | 1082461  | homeotic protein HB9 - human                                   | 0.38    |
| 659    | AF104156                              | Rattus exulans isolate huahine30 mitochondrial D-loop, partial sequence                       | 0.066   | 1002380  | (U24189) RRM-type RNA binding protein [Caenorhabditis elegans] | 0.29    |
| 660    | X97581                                | M.musculus mRNA for spalt transcription factor  | 0.066   | 4107313  | (AL035075) putative myosin heavy chain                         | 0.28    |
| 661    | D85378                                | Human close H20 N-acetylglucosaminyltransferase III DNA, exon2                                | 0.066   | 2114473  | (U96963) p140mDia [Mus musculus]                               | 0.22    |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 662                                   | M97561         | Human (clone LA 179) chromosome 21 sequence   | 0.065   | <NONE>  | <NONE>  | <NONE>  |
| 663                                   | AE001373       | Plasmodium falciparum chromosome 2. Section 10 of 73 of the complete sequence   | 0.065   | <NONE>  | <NONE>  | <NONE>  |
| 664                                   | S75479         | growth hormone receptor, growth hormone binding protein {GHR/BP gene} [mice, C57 black/6, Genomic, 179 nt, segment 8 of 10] | 0.065   | <NONE>  | <NONE>  | <NONE>  |
| 665                                   | AF032922       | Homo sapiens syntaxin 4 binding protein UNC-18c (UNC-18c) mRNA, complete cds  | 0.065   | 3061308   | (AB006074)  | 0.82    |
| 666                                   | S80986         | Svp[40]=svp-related nuclear receptor/retinoid signaling modulator [zebrafishes, mRNA, 3876 nt]                              | 0.065   | 1326288   | (U58734) weak similarity to ankyrin G [Caenorhabditis elegans]  | 0.12    |
| 667                                   | X59552         | G.domesticus mRNA for ventricular myosin heavy chain  | 0.065   | 2497098   | HYPOTHETICAL 74.2 KD PROTEIN IN AMDI-RAD52 INTERGENIC REGION<br>>gi 1077180 pir  S49745 probable membrane protein YML034w - yeast (Saccharomyces cerevisiae) >gi 575685 (Z46659) unknown orf, len: 656, CAI:0.13 [Saccharomyces cerevisiae] | 0.014   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 668    | M72980                                | Anthonomus grandis vitellogenin gene (VTG), complete cds.  | 0.065   | 3242750  | (AC005164) match to ESTs AA731149 (NID:g2140138), AA731908 (NID:g752719), AA287837 (NID:g1933519), AA262811 (NID:g1898382), and AA825820 (NID:g2899132) | 5e-33   |
| 669    | AB014546                              | Homo sapiens mRNA for KIAA0646 protein, complete cds   | 0.064   | <NONE>   | <NONE>  | <NONE>  |
| 670    | M30039                                | Sheeppox virus strain KS-1 ORF HM1 gene, partial cds; ORF HM2 and ORF HM3 genes, complete cds; and ORF HM4 gene, partial cds | 0.064   | <NONE>   | <NONE>  | <NONE>  |
| 671    | Z68013                                | Caenorhabditis elegans cosmid W02H3, complete sequence [Caenorhabditis elegans]  | 0.064   | <NONE>   | <NONE>  | <NONE>  |
| 672    | AF041332                              | Bodo saltans unknown mRNA, kinetoplast gene encoding kinetoplast protein, complete cds                                       | 0.064   | <NONE>   | <NONE>  | <NONE>  |
| 673    | <NONE>                                | <NONE>   | <NONE>  | <NONE>   | <NONE>  | <NONE>  |
| 674    | U41289                                | Dictyostelium discoideum K7 kinesin-like protein mRNA, complete cds  | 0.064   | 3482972  | (AL031369) putative protein   | 9.3     |
| 675    | M37395                                | L.lactis (strain SK11) proteinase plasmid pSK111 DNA, complete cds.  | 0.064   | 993019   | (X87246) alternative start codon [Pseudorabies virus]   | 9.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 676    | Z15030                                | H.sapiens gene for ventricular myosin light chain 2 > :: gb L01652 HUMVML C Human ventricular myosin light chain 2 gene, seven exons. | 0.064   | 730343   | PROLACTIN RECEPTOR PRECURSOR (PRL-R) mouse >gi 220576 gnl PID d1001535 (D10214) prolactin receptor precursor [Mus musculus] >gi 293770 (L14811) prolactin receptor [Mus musculus] >gi 347842 (L13593) prolactin receptor [Mus musculus] receptor:ISOTYPE=long form [Mus musculus] | 9.1     |
| 677    | Z12021                                | G.max gene for catalase   | 0.064   | 2498711  | ORIGIN RECOGNITION COMPLEX PROTEIN SUBUNIT 2 >gi 1185461 (U38472) essential ORC2-related fission replication factor Orp2 [Schizosaccharomyces pombe]  | 5.3     |
| 678    | L05668                                | Entamoeba histolytica protein serine/threonine kinase (pstk1) gene, complete cds.   | 0.064   | 733140   | (U22453) carboxypeptidase [Simulium vittatum]   | 5.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION   | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 679    | U50715                                | Mus musculus alpha-galactosidase A gene, complete cds | 0.064   | 125398   | HYGROMYCIN-B KINASE (HYGROMYCIN B PHOSPHOTRANSFERASE) (APH(7"))<br>>gi 66885 pir  WGSMEH hygromycin B phosphotransferase (EC 2.7.1.-)- Streptomyces hygroscopicus<br><gi 581682 (X03615) pot. Hyg protein [Streptomyces hygroscopicus] phosphotransferase [synthetic construct]<br>>gi 2739064 cloning vector] >gi 2739068 (AF025747) hygromycin B phosphotransferase [unidentified cloning vector] | 2.3     |
| 680    | Z28182                                | S.cerevisiae chromosome XI reading frame ORF YKL 182w | 0.064   | 1079035  | Om(2D) protein - fruit fly (Drosophila ananassae)<br>>gi 443770 gnl PID d1006095 (D26553) ORF   | 1.8     |
| 681    | M29917                                | Human ornithine aminotransferase gene, exon 1.        | 0.064   | 2317934  | (U97553) unknown [murine herpesvirus 68]  | 1.4     |
| 682    | AB020709                              | Homo sapiens mRNA for KIAA0902 protein, complete cds  | 0.064   | 861404   | (U29154) T07F12.3 gene product [Caenorhabditis elegans]   | 0.47    |
| 683    | AB014546                              | Homo sapiens mRNA for KIAA0646 protein, complete cds  | 0.064   | 1708118  | HOMEBOX PROTEIN HB9<br>>gi 507425   | 0.35    |
| 684    | AB010427                              | Homo sapiens mRNA for NORI-1, complete cds            | 0.064   | 2388676  | (AF015539) precollagen P [Mytilus edulis]   | 0.018   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 685    | U34774                                | Orf virus ankyrin-like repeat protein, F11L homolog, and F12L homolog genes, complete cds. | 0.064   | 731668   | SSF1 PROTEIN<br>>gi 626624 pir  S4670<br>0 SSF1 protein - yeast<br>(Saccharomyces cerevisiae)                                | 1e-05   |
| 686    | AF022861                              | Mus musculus neuropilin-2(a5) mRNA, alternatively spliced, complete cds                    | 0.064   | 491978   | (AF073359)<br>benzaldehyde dehydrogenase<br>[Pseudomonas sp. DJ77]   | 1e-05   |
| 687    | U14331                                | Sus scrofa myogenin gene, complete cds   | 0.064   | 2781386  | (AC004010) similar to<br>Leucinerich transmembrane<br>proteins; 44% similarity to U42767<br>(PID:g1736918)<br>[Homo sapiens] | 3e-33   |
| 688    | AF074870                              | Chironomus pallidivittatus clone 1219 non-telomeric Ssp repeat sequence                    | 0.063   | <NONE>   | <NONE>   | <NONE>  |
| 689    | Z25523                                | H.sapiens repeat region DNA.   | 0.063   | <NONE>   | <NONE>   | <NONE>  |
| 690    | AE001378                              | Plasmodium falciparum chromosome 2, section 15 of 73 of the complete sequence              | 0.063   | <NONE>   | <NONE>   | <NONE>  |
| 691    | Z72947                                | S.cerevisiae chromosome VII reading frame ORF YGR 162w                                     | 0.063   | <NONE>   | <NONE>   | <NONE>  |
| 692    | Y14723                                | Choanomphalus incertus mitochondrial cytochrome c oxidase subunit I gene, partial          | 0.063   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 693    | X74103                                | Streptomyces sp. Gene for alkaline serine protease I | 0.063   | 1730713  | HYPOTHETICAL 108.5 KD PROTEIN IN UME3-PUB1 INTERGENIC REGION<br>>gi 2131866 pir  S62935 hypothetical protein YNL023c - yeast (Saccharomyces cerevisiae)<br>>gi 1301855 gnl PID e239870 (Z71299) ORF YNL023c [Saccharomyces cerevisiae]        | 6.7     |
| 694    | AF039843                              | Homo sapiens Sprouty 2 (SPRY2) mRNA, complete cds    | 0.063   | 232217   | GLUTATHIONES-TRANSFERASE HST-6.0 (GST B1-1)<br>>gi 421198 pir  S29772 glutathione transferase (EC 2.5.1.18) - Proteus mirabilis<br>>gi 2126142 pir  S71882 glutathione transferase (EC 2.5.1.18) B - Proteus mirabilis >gi 1053076 (U38482)   | 5.2     |
| 695    | M63650                                | Mouse M-twist gene mRNA, complete cds.               | 0.063   | 1730141  | FRAGILE X MENTAL RETARDATION SYNDROME RELATED PROTEIN 2<br>>gi 2135129 pir  S60173 fragile X mental retardation syndrome related protein - human<br>>gi 1098637 (U31501) fragile X mental retardation syndrome related protein [Homo sapiens] | 1.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 696    | Y13298                                | Homo sapiens GDP dissociation inhibitor beta pseudogene      | 0.063   | 1085930  | hypothetical protein 4 - fowl adenovirus 1  | 1.3     |
| 697    | X56600                                | Rat SOD-2 gene for manganese-containing superoxide dismutase | 0.063   | 3882143  | (AB018254) KIAA0711 protein [Homo sapiens]  | 0.60    |
| 698    | Z23107                                | M.musculus mRNA for 5HTx serotonin receptor                  | 0.063   | 1708162  | HUNTINGTIN (HUNTINGTON'S DISEASE PROTEIN HOMOLOG) (HD PROTEIN)  | 0.45    |
| 699    | M20670                                | Plasmodium vivax circumsporozoite protein gene, 3' end.      | 0.063   | 4033395  | DNA GYRASE SUBUNIT B subunit [Myxococcus xanthus]   | 0.35    |
| 700    | Z62997                                | H.sapiens CpG DNA, clone 76g11, reverse read cpg 76g11.rtl.  | 0.063   | 1350911  | RETINOIC ACID RECEPTOR RXR-BETA sapiens] >gi 3172498 (AF065396) retinoic X receptor B dJ1033B10.11 (retinoid X receptor beta (RXRB)) [Homo sapiens] >gi 4249766 (AF120161) retinoic X receptor beta | 0.16    |
| 701    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds      | 0.063   | 2981200  | (AF048732) cyclin T2b [Homo sapiens]  | 0.090   |
| 702    | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds   | 0.063   | 3877951  | (Z81555) predicted using Genefinder   | 6e-07   |
| 703    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds      | 0.063   | 3393018  | (AL031174) hypothetical protein   | 2e-10   |
| 704    | D90872                                | E.coli genomic DNA, Kohara clone #419(54.7-55.1 min.)        | 0.063   | 2498198  | CYTOCHROME B561 (CYTOCHROME B-561)  | 3e-19   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 705                                   | M25528         | M.crystallinum<br>ferredoxin-NADP+<br>reductase (fnrA)<br>mRNA, complete cds.                  | 0.062   | <NONE>  | <NONE>  | <NONE>  |
| 706                                   | U45256         | Strongyloides ratti<br>microsatellite B DNA  | 0.062   | <NONE>  | <NONE>  | <NONE>  |
| 707                                   | U95102         | Xenopus laevis<br>mitotic<br>phosphoprotein 90<br>mRNA, complete cds                           | 0.062   | <NONE>  | <NONE>  | <NONE>  |
| 708                                   | AF044317       | Homo sapiens<br>TEL/AML1 fusion<br>gene, partial sequence                                      | 0.062   | <NONE>  | <NONE>  | <NONE>  |
| 709                                   | Z73975         | Caenorhabditis<br>elegans cosmid<br>T06E8, complete<br>sequence<br>[Caenorhabditis<br>elegans] | 0.062   | 3108187   | (AC004663) Notch 3<br>[Homo sapiens]  | 2.9     |
| 710                                   | X54232         | Human mRNA for<br>heparan sulfate<br>proteoglycan  | 0.062   | 1076741   | chitinase (EC 3.2.1.14)<br>precursor - rice<br>precursor - rice<br>>gi 807955 (X87109)<br>chitinase [Oryza<br>sativa] | 0.59    |
| 711                                   | X03073         | Bovine retinal mRNA<br>for transducin beta-<br>subunit   | 0.062   | 477578  | sialidase -<br>Actinomyces viscosus<br>>gi 141852   | 0.087   |
| 712                                   | Y12573         | D.melanogaster Jun<br>and 14-3-3 zeta gene   | 0.062   | 3879551   | (Z70756) similar to<br>collagen   | 0.073   |
| 713                                   | L27573         | Bombus terrestris<br>mitochondrial<br>cytochrome oxidase I,<br>partial cds.                    | 0.062   | 1684959   | (U20600) NADH<br>dehydrogenase subunit<br>[Vanda lamellata]   | 0.039   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 714    | U58994                                | Human ladinin (LAD) gene, complete cds                         | 0.062   | 2811078  | AMINOPEPTIDASE B (ARGINYL AMINOPEPTIDASE) (ARGININE AMINOPEPTIDASE) (CYTOSOL AMINOPEPTIDASE IV) (AP-B) >gi 2039143 (U61696) aminopeptidase B [Rattus norvegicus]  | 9e.06   |
| 715    | AB014553                              | Homo sapiens mRNA for KIAA0653 protein, partial cds            | 0.062   | 1326350  | (U58748) similar to potential transmembrane domains in S.cerevisiae nuclear division RFT1 protein (SP:P38206)   | 5e-10   |
| 716    | L16898                                | Mus musculus collagen alpha 1 type XVIII mRNA, 5'end.          | 0.062   | 1723657  | HYPOTHETICAL 38.5 KD PROTEIN IN ERV1-GLS2 INTERGENIC REGION >gi 2132587 pir  S6432 2 probable membrane protein YGR031w - yeast (Saccharomyces cerevisiae) >gi 1323010 gnl PID e 243277 (Z72816) ORF YGRO031w [Saccharomyces cerevisiae] | 1e-14   |
| 717    | X99343                                | M.tuberculosis guaA/B & choD genes                             | 0.062   | 3873807  | (Z49907) B0491.1 [Caenorhabditis elegans]   | 2e-19   |
| 718    | AF010193                              | Homo sapiens MAD-related gene SMAD7 (SMAD7) mRNA, complete cds | 0.061   | <NONE>   | <NONE>  | <NONE>  |
| 719    | L10182                                | Myrmeleon sp. 18S ribosomal RNA.                               | 0.061   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 720    | Y14723                                | Choanomphalus incertus mitochondrial cytochrome c oxidase subunit I gene, partial | 0.061   | <NONE>   | <NONE>  | <NONE>  |
| 721    | L27840                                | Bovine respiratory syncytial virus nucleoprotein mRNA, complete cds.              | 0.061   | 542955   | nucleoporin p62 - human   | 8.6     |
| 722    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                           | 0.061   | 494454   | Sus scrofa<br>>gi 494455 pdb 1POS <br>B Sus scrofa Sus scrofa<br>>gi 1421210 pdb 1PCP <br>Porcine Spasmolytic Protein (Psp) (Nmr, 19 Structures)<br>Spasmolytic Polypeptide<br>>gi 1633061 pdb 2PSP <br>B Chain B. Porcine Pancreatic Spasmolytic Polypeptide | 2.9     |
| 723    | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds                        | 0.061   | 3845272  | (AE001417) hypothetical protein [Plasmodium falciparum]   | 1.3     |
| 724    | U26463                                | Sporidiobolus salmonicolor NADPH-dependent aldehyde reductase gene, complete cds  | 0.061   | 1710288  | (U79302) unknown [Homo sapiens]   | 0.44    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 725    | AF035443                              | Xenopus laevis weel homolog mRNA, complete cds                                   | 0.061   | 3979720  | (????) cDNA EST EMBL:D33048 comes from this gene; cDNA EST EMBL:D35780 comes from this gene; cDNA EST yk442c6.3 comes from this gene; cDNA EST yk398f6.3 comes from this gene; cDNA E... >gi 3979816 gnl PID e1358315 EST EMBL:D35780 comes from this gene; cDNA EST yk442c6.3 comes from this gene; cDNA EST yk442c6.5 comes from this gene; cDNA EST yk398f6.3 comes from this gene; cDNA E... | 2e-04   |
| 726    | Z48548                                | Caenorhabditis elegans cosmid ZK1321, complete sequence [Caenorhabditis elegans] | 0.061   | 3183491  | HYPOTHETICAL 83.8 KD PROTEIN C27F2.7 IN CHROMOSOME III >gi 1065510 (U40419) C27F2.7 gene product [Caenorhabditis elegans]  | 3e-11   |
| 727    | X61489                                | Zea mays pep gene for (C3 type) phosphoenolpyruvate carboxylase                  | 0.061   | 2496887  | HYPOTHETICAL 32.0 KD PROTEIN C09F6.2 IN CHROMOSOME III >gi 732538 (U22832) C09F5.2 gene product [Caenorhabditis elegans]   | 1e-15   |
| 728    | AF025408                              | Drosophila melanogaster Windbeutel (wind) gene, complete cds                     | 0.061   | 3702295  | (AC005783) R33083_1 [Homo sapiens]   | 2e-60   |
| 729    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds                                       | 0.060   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 730    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 0.060   | <NONE>   | <NONE>   | <NONE>  |
| 731    | Y08682                                | H.sapiens mRNA for carnitine palmitoyltransferase I type I                           | 0.060   | 3319446  | (AF077541) contains similarity to class-I aminoacyl-tRNA synthetases [Caenorhabditis elegans]  | 8.1     |
| 732    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                              | 0.060   | 1041119  | (D78016) TRAE (Enterococcus faecalis]  | 8.1     |
| 733    | AF064030                              | Helianthus tuberosus lectin 2 mRNA, complete cds                                     | 0.060   | 632209   | regulatory protein Rex - primate T-lymphotropic virus PTLV-L (fragment)  | 3.7     |
| 734    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 0.060   | 3098348  | (AF037401) neuropeptide Y/peptide YY receptor Yc [Danio rerio]   | 2.1     |
| 735    | U95102                                | Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds                          | 0.060   | 125978   | LAR PROTEIN PRECURSOR (LEUKOCYTE ANTIGEN RELATED >gi 70146 pir  TDHUL K leukocyte antigen-related protein precursor - human >gi 34267 sapiens] | 1.2     |
| 736    | U76523                                | Sambucus nigra lectin precursor mRNA, complete cds                                   | 0.060   | 2055394  | (U87306) transmembrane receptor UNC5H2 [Rattus norvegicus]   | 0.32    |
| 737    | U69668                                | Human nuclear pore complex-associated protein TPR                                    | 0.060   | 4127854  | (Y14063) ChT1 thymocyte antigen [Gallus gallus]  | 9e-04   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 738    | AB014533                              | Homo sapiens mRNA for KIAA0654 protein, partial cds                     | 0.060   | 1326350  | (U58748) similar to potential transmembrane domains in S.cerevisiae nuclear division RFT1 protein (SP:P38206)   | 1e-09   |
| 739    | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds              | 0.060   | 2632098  | (Y15513) Prodos protein [Drosophila melanogaster]   | 5e-10   |
| 740    | Z96260                                | H.sapiens telomeric DNA sequence, clone 12QTEL101, read 12QTELOO101.seq | 0.059   | <NONE>   | <NONE>  | <NONE>  |
| 741    | M93128                                | Mouse homeobox protein (EVX2) mRNA, complete cds                        | 0.059   | <NONE>   | <NONE>  | <NONE>  |
| 742    | AB012106                              | Brassica rapa mRNA for SRK45,complete cds                               | 0.059   | 1652318  | (D90904) lysostaphin [Synechocystis sp.]  | 4.7     |
| 743    | AB007920                              | Homo sapiens mRNA for KIAA0451 protein, complete cds                    | 0.059   | 479491   | transcription factor brn-3b - human   | 0.71    |
| 744    | M60445                                | Human histidine decarboxylase (HDC) mRNA, complete cds                  | 0.058   | <NONE>   | <NONE>  | <NONE>  |
| 745    | U01836                                | Ustilago maydis exodeoxyribonuclease (REC1) gene, complete cds.         | 0.058   | 1171908  | OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPS >gi 1075086 pir  D64184 oligopeptide transport system permease protein (oppC)C homolog - Haemophilus influenzae (strain RdKW20) permease protein (oppC) [Haemophilus influenzae Rd] | 1.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 746    | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 0.058   | 3193265  | (AF069131) chitinase [Bacillus subtilis]   | 0.002   |
| 747    | AB012105                              | Brassica rapa mRNA for SLG45, complete cds  | 0.057   | 43385  | (U03978) dynein heavy chain isotype 7A [Tripneustes gratilla]  | 3.4     |
| 748    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme  | 0.056   | <NONE>   | <NONE>   | <NONE>  |
| 749    | Y16828                                | Lagopus lagopus genomic microsatellite sequence, LLST4  | 0.056   | 3328578  | (AE001299) hypothetical protein [Chlamydia trachomatis]  | 4.3     |
| 750    | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                             | 0.055   | <NONE>   | <NONE>   | <NONE>  |
| 751    | AF074385                              | Sambucus nigra hevein-like protein mRNA, complete cds   | 0.055   | 137339   | 69 KD PROTEIN >gi 94375 pir  S19150 hypothetical protein, 69K - turnip yellow mosaic virus               | 0.69    |
| 752    | U76523                                | Sambucus nigra lectin precursor mRNA, complete cds  | 0.035   | <NONE>   | <NONE>   | <NONE>  |
| 753    | M92069                                | Human retrovirus-like sequence-isoleucine c   | 0.034   | <NONE>   | <NONE>   | <NONE>  |
| 754    | S78516                                | GIL=ankyrin-like repeat [orf virus OV, NZ2, Genomic, 1608 nt]   | 0.033   | 2804465  | (AF043700) contains similarity to human RNA-binding protein FUS/TLS (SW:Q28009) [Caenorhabditis elegans] | 0.15    |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 755                                   | M15646         | Chicken myosin alkali light chain mRNA, complete cds, clone pF1.                     | 0.027   | 3334221   | 4-HYDROXYPHENYL PYRUVATE DIOXYGENASE 4-hydroxyphenylpyruvate dioxygenase [Mycosphaerella graminicola]   | 6e-17   |
| 756                                   | AF027174       | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 0.025   | 3877815   | (Z96048) predicted using Genefinder   | 5.0     |
| 757                                   | AJ002291       | Streptococcus pneumoniae pbp1b gene, complete  | 0.025   | 3880487   | (Z6814) similar to ribosephosphate pyrophosphokinase; cDNA EST EMBL:D73173 comes from this gene; cDNA EST EMBL:D70909 comes from this gene; cDNA EST EMBL:D73449 comes from this gene; cDNA EST EMBL:D76167 comes from this ge... | 1.7     |
| 758                                   | X79104         | C.botulinum (NCTC 7272 type A) HA-33 and P-2 genes.                                  | 0.024   | 2648615   | (AE000970) tungsten formylmethanofuran dehydrogenase, subunit B (fwdB2) [Archaeoglobus fulgidus]  | 6.1     |
| 759                                   | U95102         | Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds                          | 0.024   | 1663698   | (D83785) expressed ubiquitously; product similar to D.melanogaster mam protein. [Homo sapiens]  | 4.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 760    | U36197                                | Chlamydomonas reinhardtii cobalamin-independent methionine synthase mRNA, complete cds  | 0.024   | 585723   | PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR GAMMA (PPAR-GAMMA) >gi 283818 pir C42214 peroxisome proliferator-activated receptor gamma chain - African clawed frog >gi 214668 (M84163) peroxisome proliferator activated receptor gamma [Xenopus laevis] | 0.42    |
| 761    | L38865                                | Macaca mulatta (close MMVA63) T-cell receptor alpha (TCR A) mRNA, partial cds.          | 0.023   | <NONE>   | <NONE>   | <NONE>  |
| 762    | AF035948                              | Mus musculus insulin receptor substrate-3   | 0.023   | 2500587  | SPLICEOSOME ASSOCIATED PROTEIN 49 spliceosome-associated protein SAP-49 - human >gi 556217   | 0.40    |
| 763    | X98890                                | S.tuberosum mRNA for inorganic phosphate transporter, StPT1                             | 0.023   | 110072   | proline-rich protein MP4 - mouse >gi 53182   | 0.18    |
| 764    | X91212                                | L.esculentum mRNA for HD-ZIP protein  | 0.022   | <NONE>   | <NONE>   | <NONE>  |
| 765    | AC004498                              | Homo sapiens chromosome 5, P1 clone 1209C1 (LBNLH104), complete sequence [Homo sapiens] | 0.022   | <NONE>   | <NONE>   | <NONE>  |
| 766    | U07083                                | Human prostatic acid phosphatase (ACPP) gene, exon 1                                    | 0.022   | <NONE>   | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 767                                   | X98890         | S.tuberosum mRNA for inorganic phosphate transporter, StPT1                    | 0.022   | <NONE>  | <NONE>  | <NONE>  |
| 768                                   | X56488         | L.esculentum LAT59 gene 5' flanking region, expressed during pollen maturation | 0.022   | <NONE>  | <NONE>  | <NONE>  |
| 769                                   | M34651         | Pseudorabies virus with upstream and downstream sequences.                     | 0.022   | <NONE>  | <NONE>  | <NONE>  |
| 770                                   | X66727         | P.taeda gene for protochlorophyllide reductase                                 | 0.022   | 3878517   | (Z92806) K10G4.4 [Caenorhabditis elegans]   | 4.3     |
| 771                                   | U95102         | Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds                    | 0.022   | 1854452   | (D89501) similar to salivary proline-rich protein P-B [Homo sapiens]                    | 4.3     |
| 772                                   | U95098         | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds                     | 0.022   | 3021699   | (AB005298) BAI 2 [Homo sapiens]   | 0.64    |
| 773                                   | X71932         | H.sapiens XB gene for tenascin-X intron 14                                     | 0.022   | 627059  | liver stage antigen LSA-1 - Plasmodium falciparum >gi 9916 (X56203) liver stage antigen | 0.058   |
| 774                                   | X87369         | C.perfringens nanH gene & ORF1.2.3 & 4   | 0.022   | 2062407   | (U78975) poly(ADP-ribose) glycohydrolase [Bos taurus]                                   | 0.056   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 775    | Y14971                                | Gallus gallus mRNA for K60 protein   | 0.022   | 134091   | UI SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KD (U1 SNRNP 70 KD) >gi 85864 pir  S02016 U1 snRNP 70K protein - African clawed frog >gi 65179 (X12430) U1 70K [Xenopus laevis] | 0.032   |
| 776    | AF003133                              | Caenorhabditis elegans cosmid T21E3  | 0.022   | 1709997  | DNA REPAIR PROTEIN RAD18 >gi 1150622 protein rad18 [Schizosaccharomyces pombe]  | 2e-08   |
| 777    | AF003133                              | Caenorhabditis elegans cosmid T21E3  | 0.022   | 1709997  | DNA REPAIR PROTEIN RAD18 >gi 1150622 protein rad18 [Schizosaccharomyces pombe]  | 2e-08   |
| 778    | U57645                                | Human helix-loop-helix proteins Id-1 (ID-1) and Id-1' (ID-1) genes, complete cds | 0.021   | <NONE>   | <NONE>  | <NONE>  |
| 779    | U67570                                | Methanococcus jannaschii section 112 of 150 of the complete genome               | 0.021   | <NONE>   | <NONE>  | <NONE>  |
| 780    | L01584                                | Trypanosoma cruzi calcium-binding protein (CUB2.8) gene, complete cds.           | 0.021   | <NONE>   | <NONE>  | <NONE>  |
| 781    | L04787                                | Borrelia hermsii outer membrane lipoprotein                                      | 0.021   | <NONE>   | <NONE>  | <NONE>  |
| 782    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                          | 0.021   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION   | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 783    | L36890                                | Saccharomyces cerevisiae mitochondrion transfer RNA-Thr1 (tRNA-Thr) gene; transfer RNA-Val (tRNA-Val) gene; oxi2 gene, complete cds; ORF2 and origin of replication (ori5). | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 784    | M76741                                | Homo sapiens biliary glycoprotein (BGP) gene, partial cds.  | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 785    | M87504                                | Tetrahymena thermophila histone H3 (HHT2) gene, complete cds  | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 786    | U94346                                | Human calpain-like protease (htra-3) mRNA, complete cds   | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 787    | L01584                                | Trypanosoma cruzi calcium-binding protein (CUB2.8) gene, complete cds   | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 788    | U36530                                | Pongo pygmaeus CT microsatellite, clone #1, from the tandemly repeated genes encoding U2 small nuclear RNA (RNU2 locus)   | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 789    | X03833                                | Human gene for interleukin 1 alpha (IL-1 alpha)   | 0.021   | 416974   | EARLY TRANSCRIPTION FACTOR 70 KD SUBUNIT               | 8.9     |
| 790    | U20806                                | Dictyostelium discoideum guanine nucleotide-binding protein alpha subunit 5 (G alpha 5) gene, complete cds.   | 0.021   | 1401211  | (U58510) RNA helicase homolog [Chlorarachnion CCMP621] | 8.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 791    | Z59258                                | H.sapiens CpG DNA, clone 13d2, reverse read cpg13d2.rtlc.  | 0.021   | 3121732  | ACONITATE HYDRATASE (CITRATE HYDRO-LYASE) (ACONITASE) >gi 2183256 (AF002133) aconitase [Mycobacterium avium]   | 7.0     |
| 792    | AF030692                              | Plasmodium falciparum strain 7G8 chloroquine resistance candidate protein (cg2) gene, complete cds | 0.021   | 3024190  | NINE PROTEIN >gi 2120251 pir  S665 81 hypothetical protein 56 - phage 82 >gi 1051114 (X92588) orf56; related to nin60 (ninE) of bacteriophage lambda | 5.8     |
| 793    | U67570                                | Methanococcus jannaschii section 112 of 150 of the complete genome                                 | 0.021   | 2341037  | (AC000104) F19P19.17 [Arabidopsis thaliana]  | 4.0     |
| 794    | D86566                                | Human DNA for NOTCH4, partial cds  | 0.021   | 1708619  | NUCLEAR FACTOR NF-KAPPA-B P100 SUBUNIT (H2TF1) (ONCOGENE LYT-10) (LYT10) [CONTAINS: NUCLEAR FACTOR NF-KAPPA-B P52 SUBUNIT]                           | 3.1     |
| 795    | L11648                                | Streptomyces coelicolor sigma factor (rpoX) gene, complete cds.                                    | 0.021   | 79833  | Hypothetical 119.5K protein (uvrA region) - Micrococcus luteus   | 1.8     |
| 796    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds  | 0.021   | 128000   | NEUROENDOCRINE CONVERTASE 1 PRECURSOR (NEC 1) PC1) (PROHORMONE CONVERTASE 1) propeptide processing protease [Mus cookii]                             | 1.0     |



| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 797                                   | U30938         | Rattus norvegicus microtubule-associated protein 2                            | 0.021   | 468600  | (X74416) beta-3 integrin [Takifugu rubripes]  | 1.0     |
| 798                                   | D82364         | Chicken mRNA for TSC-22 variant, complete cds, clone SLFEST52                 | 0.021   | 693723  | 27 kda amelogenin {alternatively spliced}   | 0.61    |
| 799                                   | U40041         | Gallus gallus eHAND mRNA, complete cds  | 0.021   | 3449308   | (AB011541) MEGF8 [Homo sapiens]   | 0.21    |
| 800                                   | X71932         | H.sapiens XB gene for tenascin-X, intron 14                                   | 0.021   | 627059  | Liver stage antigen LSA-1-Plasmodium falciparum >gi 9916 (X56203) liver stage antigen                               | 0.054   |
| 801                                   | AF042333       | Oryza sativa 24-methylene lophenol C24(1)methyltransferase mRNA, complete cds | 0.021   | 854065  | (X83413) U88 [Human herpesvirus 6]  | 0.014   |
| 802                                   | L37380         | Rat apical endosomal glycoprotein mRNA, complete cds.                         | 0.021   | 3334377   | TRANSMEMBRANE PROTEASE, SERINE 2  | 1e-05   |
| 803                                   | AF003133       | Caenorhabditis elegans cosmid T21E3   | 0.021   | 1709997   | DNA REPAIR PROTEIN RAD18>gi 1150622 protein rad18 [Schizosaccharomyces pombe]                                       | 3e-08   |
| 804                                   | X57689         | Rabbit mRNA for calcium channel BI-2 (lambda CBP109 and CB101)                | 0.021   | 2959370   | (AL022117) hypothetical protein   | 1e-10   |
| 805                                   | U95102         | Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds                   | 0.021   | 1109830   | (U41534) coded for by C. elegans cDNA CEESI42F; Similar to helicases of SNF2/RAD54 family. [Caenorhabditis elegans] | 5e-11   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 806    | X77753                                | H. sapiens TROP-2 gene  | 0.021   | 1723657  | HYPOTHETICAL 38.5 KD PROTEIN IN ERV1-GLS2 INTERGENIC REGION<br>>gi 2132587 pir  S64322 probable membrane protein YGR031w - yeast (Saccharomyces cerevisiae)<br>>gi 1323010 gnl PID e243277 (Z72816) ORF YGR031w [Saccharomyces cerevisiae] | 5e-11   |
| 807    | X98890                                | S. tuberosum mRNA for inorganic phosphate transporter, StPT1  | 0.021   | 2137872  | zinc finger protein PZF - mouse<br>>gi 453376  | 2e-19   |
| 808    | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds  | 0.020   | <NONE>   | <NONE>   | <NONE>  |
| 809    | AJ224935                              | Homo sapiens Promoter Region and PCK2 gene  | 0.020   | <NONE>   | <NONE>   | <NONE>  |
| 810    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds   | 0.020   | <NONE>   | <NONE>   | <NONE>  |
| 811    | X99941                                | A. thaliana GBF1 gene   | 0.020   | <NONE>   | <NONE>   | <NONE>  |
| 812    | X65138                                | M. musculus mRNA for tyrosine kinase >::gb S57168 S57168 Sek=Eph-related receptor protein tyrosine kinase [mice, mRNA, 4242 nt] | 0.020   | <NONE>   | <NONE>   | <NONE>  |
| 813    | L04787                                | Borrelia hermsii outer membrane lipoprotein   | 0.020   | <NONE>   | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 814                                   | AJ223633       | Enterococcus faecium genes encoding enterocin L50A and enterocin L50B plus 5' and 3' flanking regions | 0.020   | <NONE>  | <NONE>   | <NONE>  |
| 815                                   | AB012106       | Brassica rapa mRNA for SRK45, complete cds  | 0.020   | <NONE>  | <NONE>   | <NONE>  |
| 816                                   | AE001539       | Helicobacter pylori, strain J99 section 100 of 132 of the complete genome                             | 0.020   | 172292  | (L11895) transmembrane protein [Saccharomyces cerevisiae]  | 8.4     |
| 817                                   | AF074386       | Sambucus nigra hevein-like protein mRNA, complete cds   | 0.020   | 94173   | pol polyprotein - Chinese hamster intracisternal A-particle CHIAP34  | 8.0     |
| 818                                   | M55264         | Herpesvirus saimiri dihydrofolate reductase (DHFR) and snRNA (HSUR) genes, complete cds.              | 0.020   | 2924250   | (Z98745) dJ29K1.2 [Homo sapiens]   | 6.5     |
| 819                                   | AF052163       | Homo sapiens clone 24456 mRNA sequence  | 0.020   | 1706288   | D(4) DOPAMINE RECEPTOR (D(2C) DOPAMINE RECEPTOR)<br>>gi 2119482 pir  I49246 D4 dopamine receptor - mouse<br>>gi 758427 (U19880) D4 dopamine receptor [Mus musculus]<br>>gi 1095539 prf  2109259A dopamine D4 receptor [Mus musculus] | 4.9     |
| 820                                   | AF074387       | Sambucus nigra hevein-like protein mRNA, complete cds   | 0.020   | 2113798   | (Z83259) AmphiBrf38 [Branchiostoma floridae]   | 4.7     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 821                                   | AF052163       | Homo sapiens clone 24456 mRNA sequence                                    | 0.020   | 3874733   | (Z67754) cDNA EST EMBL:T02354 comes from this gene; cDNA EST EMBL:D32698 comes from this gene; cDNA EST EMBL:D35411 comes from this gene  | 4.7     |
| 822                                   | L11002         | Rat ankyrin binding glycoprotein-1 related mRNA sequence.                 | 0.020   | 552132  | (K01664) Bkm-like protein [Drosophila melanogaster]   | 3.8     |
| 823                                   | AE001539       | Helicobacter pylori, strain J99 section 100 of 132 of the complete genome | 0.020   | 172292  | (L11895) transmembrane protein [Saccharomyces cerevisiae]   | 3.8     |
| 824                                   | X98890         | S. tuberosum mRNA for inorganic phosphate transporter, StPT1              | 0.020   | 3879798   | (Z81120) Similar to TPR Domain (2 domains); cDNA EST yk390b10.3 comes from this gene; cDNA EST EMBL:D71652 comes from this gene; cDNA EST yk275f8.3 comes from this gene; cDNA EST yk393b9.3 comes from this gene; cDNA EST yk37...>gi 3880220 gnl PID e 1349842 yk390b10.3 comes from this gene; cDNA EST EMBL:D71652 comes from this gene; cDNA EST yk275f8.3 comes from this gene; cDNA EST yk393b9.3 comes from this gene; cDNA EST yk37... | 1.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 825    | U97519                                | Homo sapiens podocalyxin-like protein mRNA, complete cds  | 0.020   | 1345633  | C-1-TETRAHYDROFOLATE SYNTHASE, CYTOPLASMIC (C1-THF SYNTHASE) (METHYLENETETRAHYDROFOLATE DEHYDROGENASE / METHENYL TETRAHYDROFOLATE CYCLOHYDROLASE C1-tetrahydrofolate synthase [Rattus norvegicus] | 0.066   |
| 826    | AF003133                              | Caenorhabditis elegans cosmid T21E3   | 0.020   | 1709997  | DNA REPAIR PROTEIN RAD18 >gi 1150622 protein rad 18 [Schizosaccharomyces pombe]   | 2e-07   |
| 827    | U32857                                | Saccharomyces cerevisiae VAR1 gene, mitochondrial gene encoding mitochondrial protein, 3' processing site, partial sequence | 0.019   | <NONE>   | <NONE>  | <NONE>  |
| 828    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds  | 0.019   | 2506381  | NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING PROTEIN INT-3) mammary gene mRNA, complete cds., gene product [Mus musculus]   | 3.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 829    | AF034099                              | Laccaria bicolor glyoxal malate synthase protein mRNA, complete cds   | 0.019   | 3880930  | (AL021481) similar to Phosphoglucomutase and phosphomannomutase phosphoserine; cDNA EST EMBL:D36168 comes from this gene; cDNA EST EMBL:D70697 comes from this gene; cDNA EST yk373h9.5 comes from this gene; cDNA EST EMBL:T008... | 6e-15   |
| 830    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds.   | 0.018   | <NONE>   | <NONE>  | <NONE>  |
| 831    | U24578                                | Human RP1 and complement C4B precursor (C4B) genes, partial cds   | 0.013   | 478673   | proline-rich protein precursor - kidney bean vulgaris]  | 3.1     |
| 832    | U76523                                | Sambucus nigra lectin precursor mRNA, complete cds  | 0.011   | <NONE>   | <NONE>  | <NONE>  |
| 833    | U57649                                | Dibenzofuran-degrading bacterium DPO360 2,3-dihydroxybiphenyl 1,2-dioxygenase (bphC) gene, complete cds and 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid hydrolase | 0.011   | <NONE>   | <NONE>  | <NONE>  |
| 834    | X15642                                | Z.mays gene for phosphoenolpyruvate carboxylase   | 0.011   | <NONE>   | <NONE>  | <NONE>  |
| 835    | X51623                                | C.elegans collagen gene col-13  | 0.010   | 1695686  | (D83706) pyruvate carboxylase [Bacillus stearothermophilus]   | 3.1     |
| 836    | U83656                                | Rattus norvegicus NF-KB gene, promotor region   | 0.008   | 4240195  | (AB020660) KIAA0853 protein [Homo sapiens]  | 10.0    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 837    | AJ222657                              | Homo sapiens gene encoding retina-specific guanylyl cyclase               | 0.008   | 417704   | POL POLYPROTEIN (ORF1A/1B) [CONTAINS: RNA-DIRECTED RNA POLYMERASE; HELICASE; PROTEASE]   | 7.4     |
| 838    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds                                | 0.008   | 544024   | CHLORIDE CHANNEL PROTEIN, SKELETAL MUSCLE (CHLORIDE CHANNEL PROTEIN 1) (CLC-1) human >gi 397143 (Z25587) human C1C-1 muscle chloride channel [Homo sapiens] >gi 398161 (Z25884) human C1C-1 muscle chloride channel [Homo sapiens] | 4.6     |
| 839    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 0.008   | 532468   | (U13643) similar to reverse transcriptase; possible pseudogene [Caenorhabditis elegans]  | 3.8     |
| 840    | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 0.008   | 4101160  | (AF002589) cytochrome oxidase I [Austrofundulus limnaeus]  | 2.7     |
| 841    | AF074385                              | Sambucus nigra hevein-like protein mRNA, complete cds                     | 0.008   | 1711520  | SRB-8/9 PROTEIN >gi 1334996  | 1.6     |
| 842    | U48734                                | Human non-muscle alpha-actinin mRNA, complete cds                         | 0.008   | 2829922  | (AC002291) extensin [Arabidopsis thaliana]   | 0.11    |
| 843    | U66669                                | Homo sapiens 3-hydroxyisobutyryl-coenzyme A hydrolase mRNA, complete cds  | 0.007   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION | P VALUE |
| 844    | D16492                                | Mouse mRNA for P100 serine protease of Ra-reactive factor (RaRF), complete cds                       | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 845    | D90923                                | Human immunodeficiency virus type 1 proviral DNA for envelope glycoprotein, partial cds, isolate 03S | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 846    | AB011087                              | Homo sapiens mRNA for KIAA0515 protein, partial cds  | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 847    | AE000688                              | Aquifex aeolicus section 20 of 109 of the complete genome  | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 848    | X63723                                | B.bovis WC1.1 mRNA   | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 849    | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds  | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 850    | J00097                                | Human beta globin region Alu repetitive sequence type T.   | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 851    | D90923                                | Human immunodeficiency virus type 1 proviral DNA for envelope glycoprotein, partial cds, isolate 03S | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 852    | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds  | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 853    | X91618                                | T.castaneum hunchback gene   | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 854    | X03838                                | Rat nontranscribed spacer (NTS) downstream of 28S rRNA gene  | 0.007   | <NONE>   | <NONE>      | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION | P VALUE |
| 855    | M55049                                | Rattus norvegicus interleukin-2 receptor alpha chain (CD25) mRNA, complete cds.              | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 856    | Z64318                                | H.sapiens CpG DNA, clone 9e2, reverse read cpg9e2.rtl.a.                                     | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 857    | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds         | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 858    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds         | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 859    | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                    | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 860    | X95276                                | P.falciparum complete gene map of plastid-like DNA   | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 861    | U72396                                | Lycopersicon esculentum class II small heat shock protein Le-HSP17.6 mRNA, complete cds      | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 862    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 863    | AB000383                              | Leucania seperata nuclear polyhedrosis virus DNA for p13, xe, envelope protein, complete cds | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 864    | D86566                                | Human DNA for NOTCH4, partial cds  | 0.007   | <NONE>   | <NONE>      | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 865                                   | U76524         | Sambucus nigra<br>ribosome inactivating<br>protein precursor<br>mRNA, complete cds                | 0.007   | <NONE>  | <NONE>   | <NONE>  |
| 866                                   | AF027173       | Arabidopsis thaliana<br>cellulose synthase<br>catalytic subunit (Ath-<br>A) mRNA, complete<br>cds | 0.007   | 3047072   | (AF058825) No<br>definition line found<br>[Arabidopsis thaliana]   | 8.9     |
| 867                                   | AF027174       | Arabidopsis thaliana<br>cellulose synthase<br>catalytic subunit (Ath-<br>B) mRNA, complete<br>cds | 0.007   | 975754  | (U29359) SpaO<br>[Salmonella enterica]   | 8.6     |
| 868                                   | U76524         | Sambucus nigra<br>ribosome inactivating<br>protein precursor<br>mRNA, complete cds                | 0.007   | 1213557   | (U50199) coded for by<br>C. elegans cDNA<br>yk89e9.5; coded for<br>by C. elegans cDNA<br>cm7g5; coded for by<br>C. elegans cDNA<br>cm14b9; coded for by<br>C. elegans cDNA<br>yk52g5.5; coded for<br>by C. elegans cDNA<br>yk76e5.5; coded for<br>by C. elegans cDNA<br>yk131f11.5; c... | 8.4     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 869                                   | AB012106       | Brassica rapa mRNA<br>for SRK45, complete<br>cds                | 0.007   | 2499568   | PROTEIN-L-<br>ISOASPARTATE(D-<br>ASPARTATE) O-<br>METHYLTRANSFER<br>ASE (PROTEIN-<br>BETA-ASPARTATE<br>METHYLTRANSFER<br>ASE) PIMT)<br>(PROTEIN L-<br>ISOASPARTYL/D-<br>ASPARTYL<br>METHYLTRANSFER<br>ASE)<br>methyltransferase<br>[Drosophila<br>melanogaster]<br>>gi 1171337<br>melanogaster] | 8.3     |
| 870                                   | AF093268       | Rattus norvegicus<br>homer-1c mRNA,<br>complete cds             | 0.007   | 4092077   | (AF095353) toll-like<br>receptor 4 mutant<br>[Mus musculus]   | 6.2     |
| 871                                   | AF074386       | Sambucus nigra<br>hevein-like protein<br>mRNA, complete cds     | 0.007   | 151377  | (M80653) tetraheme<br>[Pseudomonas<br>stutzeri]   | 6.2     |
| 872                                   | L42319         | Bos taurus (clone<br>Sal3.8) tristetraprolin                    | 0.007   | 2507337   | TRANSCRIPTION<br>TERMINATION<br>FACTOR RHO  | 5.5     |
| 873                                   | M59815         | Human complement<br>component C4A gene,<br>exons 10 through 41. | 0.007   | 3876769   | (Z69637) Similarity to<br>Human Prolyl 4-<br>hydroxylase alpha<br>subunit<br>(SW:P4HA_HUMAN)<br>; cDNA EST<br>yk219g12.5 comes<br>from this gene; cDNA<br>EST yk319d8.5 comes<br>from this gene; cDNA<br>EST yk339d11.5<br>comes from this gene;<br>cDNA EST<br>yk371c9.3...                    | 5.3     |
| 874                                   | X63723         | B.bovis WC1.1<br>mRNA   | 0.007   | 2969893   | (AJ001858) human<br>SIM2 [Homo sapiens]   | 5.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 875    | AB009864                              | Expression vector pME18S-FL3, complete sequence                           | 0.007   | 2137618  | p45 NF-E2 related factor 2 - mouse musculus]   | 5.1     |
| 876    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 0.007   | 2804497  | (AF043705) contains similarity to C2H2-type zinc fingers   | 5.0     |
| 877    | U95102                                | Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds               | 0.007   | 440298   | (L27469) product of alternative splicing [Drosophila melanogaster]   | 4.7     |
| 878    | X58869                                | Chicken mRNA for aldehyde dehydrogenase                                   | 0.007   | 1185062  | (L75945) flagellar export protein [Borrelia burgdorferi]   | 4.1     |
| 879    | AF027735                              | Nephila clavipes minor ampullate silk protein MiSp1 mRNA, partial cds     | 0.007   | 2394390  | (AF017434) pmi-like gene product [Methylobacterium extorquens]   | 4.0     |
| 880    | AF105228                              | Bos taurus tuftelin mRNA, complete cds                                    | 0.007   | 3036802  | (AL022373) putative protein  | 3.9     |
| 881    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                  | 0.007   | 2500814  | HYPOTHETICAL 60.2 KD PROTEIN T27F2.1 IN CHROMOSOME V >gi 3880311 gnl PID e 1349855 BX42 (SW:BX42_DROME); cDNA EST EMBL:C07233 comes from this gene; cDNA EST EMBL:C08532 comes from this gene; cDNA EST yk501h10.3 comes from this gene; cDNA EST yk501f1.3... | 3.8     |
| 882    | X93567                                | L.major mRNA for beta-tubulin (1404bp)                                    | 0.007   | 2317862  | (U78289) ty lactone synthase modules 4 & 5 [Streptomyces fradiae]  | 3.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 883    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds                                       | 0.007   | 3881103  | (AL032646) predicted using Genefinder; cDNA EST EMBL:D76407 comes from this gene; cDNA EST EMBL:C08999 comes from this gene; cDNA EST yk199b12.5 comes from this gene; cDNA EST yk282a4.5 comes from this gene; cDNA EST EMBL:CO... | 2.7     |
| 884    | AF041056                              | Homo sapiens WSCR4 gene, exons 3 and 4   | 0.007   | 135817   | THROMBIN RECEPTOR PRECURSOR human >gi 339677 (M62424) thrombin receptor [Homo sapiens]  | 2.2     |
| 885    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds                                    | 0.007   | 1723518  | HYPOTHETICAL 32.2 KD PROTEIN C22E12.04 IN CHROMOSOME I >gi 1220279 (Z70043) unknown   | 2.1     |
| 886    | M74798                                | Hevea brasiliensis 3-hydroxy-3-methylglutaryl-coenzyme A reductase gene, 3' end. | 0.007   | 1001282  | (D64003) poly A polymerase  | 1.9     |
| 887    | Z62997                                | H.sapiens CpG DNA, clone 76g11, reverse read cpg76g11.rta.                       | 0.007   | 1176532  | HYPOTHETICAL 111.9 KD PROTEIN C34E10.8 IN CHROMOSOME III >gi 500731 (U10402) weakly similar to protein C kinase substrate [Caenorhabditis]  | 1.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 888    | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds                                | 0.007   | 2498317  | DVA-1 POLYPROTEIN PRECURSOR<br>nematode polypeptide antigen precursor<br>[Dictyocaulus viviparus]<br>>gi 1585421 prf 2124414A polypeptide antigen/allergen<br>[Dictyocaulus viviparus] | 1.2     |
| 889    | L29426                                | Synechocystis species (strain PCC 6803) drgA gene, complete cds.                     | 0.007   | 3882275  | (AB018320) KIAA0777 protein [Homo sapiens]   | 1.1     |
| 890    | D83329                                | Mus musculus DNA for prostaglandin D2 synthase, complete cds                         | 0.007   | 1001741  | (D64004) hypothetical protein  | 0.97    |
| 891    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds   | 0.007   | 1723928  | HYPOTHETICAL 11.6 KD PROTEIN IN NUT1-ARO2 INTERGENIC REGION PRECURSOR YGL149w - yeast (Saccharomyces   | 0.94    |
| 892    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 0.007   | 121452   | GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT 12 PRECURSOR<br>>gi 82606 pir A24266 glutenin high molecular weight chain 12 precursor - wheat >gi 21779                                       | 0.79    |
| 893    | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 0.007   | 927287   | (U30294) ORF2 [Prevotella rumenicola]  | 0.35    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 894    | Y11918                                | H.sapiens IMAGE cDNA clone 26881   | 0.007   | 1055188  | (U40061) contains similarity to transmembrane domains like those found in sugar transporter proteins          | 0.26    |
| 895    | L36827                                | Mus musculus alphaA-crystallin-binding protein I                                 | 0.007   | 4063019  | (AF083061) ABC transporter TliF [Pseudomonas fluorescens]   | 0.21    |
| 896    | L36827                                | Mus musculus alphaA-crystallin-binding protein I                                 | 0.007   | 4063019  | (AF083061) ABC transporter TliF [Pseudomonas fluorescens]   | 0.20    |
| 897    | Z65719                                | H.sapiens CpG DNA, clone 54c10, reverse read cpg54c10.rfla.                      | 0.007   | 1097307  | HIC-1 gene [Homo sapiens]   | 0.20    |
| 898    | AF064029                              | Helianthus tuberosus lectin 1 mRNA, complete cds                                 | 0.007   | 1174915  | UTROPHIN (DYSTROPHIN-RELATED PROTEIN 1) (DRP1) (DRP) >gi 284488 pir  S2838 1 utrophin protein) [Homo sapiens] | 0.002   |
| 899    | AF051730                              | Mus musculus cathepsin S (CatS) gene, exon 6                                     | 0.007   | 1707017  | (U78721) RNA helicase isolog [Arabidopsis thaliana]   | 0.001   |
| 900    | U62398                                | Oryctolagus cuniculus gp42/basigin/OX-47/HT7 mRNA, complete cds.                 | 0.007   | 2370494  | (Z98944) hypothetical protein   | 2e-04   |
| 901    | X76341                                | M.musculus glutathione reductase mRNA.   | 0.007   | 3513303  | (AC005594) R26984_1 [Homo sapiens]  | 8e-07   |
| 902    | M26215                                | Rat (lambda 20B0.5) M-type 6-phosphofructo-2-kinase/fructose-2, 6-bisphosphatase | 0.007   | 3036809  | (AL022373) putative protein   | 6e-15   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION   | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 903    | AB007902                              | Homo sapiens KIAA0442 mRNA, partial cds   | 0.007   | 2662165  | (AB007902) HH0712 cDNA clone for KIAA0442 has a 574-bp insertion at position 1474 of the sequence of KIAA0442. [Homo sapiens] | 2e-17   |
| 904    | U93364                                | Lactococcus lactis cremoris plasmid pNZ4000 insertion sequence IS982 putative transposase gene and eps gene cluster (epsRXABCDEFGHKL), complete cds | 0.007   | 2731377  | (U28739) similar to alcohol dehydrogenase/ribitol dehydrogenase [Caenorhabditis elegans]                                      | 1e-31   |
| 905    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds   | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 906    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 907    | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds   | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 908    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds  | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 909    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme  | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 910    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds  | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 911    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds   | 0.006   | <NONE>   | <NONE>  | <NONE>  |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 912                                   | AF093268       | Rattus norvegicus<br>homer-1c mRNA,<br>complete cds                                | 0.006   | <NONE>  | <NONE>   | <NONE>  |
| 913                                   | AB012106       | Brassica rapa mRNA<br>for SRK45, complete<br>cds                                   | 0.006   | <NONE>  | <NONE>   | <NONE>  |
| 914                                   | AF064029       | Helianthus tuberosus<br>lectin 1 mRNA,<br>complete cds                             | 0.006   | <NONE>  | <NONE>   | <NONE>  |
| 915                                   | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds                                     | 0.006   | <NONE>  | <NONE>   | <NONE>  |
| 916                                   | AF093268       | Rattus norvegicus<br>homer-1c mRNA,<br>complete cds                                | 0.006   | 4049856   | (AF063866) ORF<br>MSV064 hypothetical<br>protein [Melanoplus<br>sanguinipes<br>entomopoxvirus]   | 9.6     |
| 917                                   | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds                                     | 0.006   | 3880536   | (Z82070) predicted<br>using Genefinder;<br>similar to Lectin C-<br>type domain short and<br>long forms (2<br>domains); cDNA EST<br>EMBL:C10633 comes<br>from this gene; cDNA<br>EST EMBL:C12424<br>comes from this gene;<br>cDNA EST yk191e7.3<br>comes from this... | 7.9     |
| 918                                   | AF012899       | Sambucus nigra<br>ribosome inactivating<br>protein precursor<br>mRNA, complete cds | 0.006   | 3877761   | (Z81552) F56G4.1<br>[Caenorhabditis<br>elegans]<br>>gi 3878615 gnl PID e<br>1348240 (Z83118)<br>F56G4.1  | 7.5     |
| 919                                   | X80289         | H.sapiens PTPL1<br>mRNA for protein<br>tyrosine phosphatase                        | 0.006   | 1168791   | CATHEPSIN E<br>PRECURSOR<br>precursor - rabbit<br>>gi 402729 (L08418)<br>procathepsin E  | 7.4     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 920                                   | AF074386       | Sambucus nigra<br>hevein-like protein<br>mRNA, complete cds   | 0.006   | 1346371   | DIACYLGLYCEROL<br>KINASE, BETA<br>DIACYLGLYCEROL<br>KINASE)<br>>gi 477059 pir  A4774<br>4 diacylglycerol<br>kinase (EC2.7.1.107)<br>beta - rat 90kDa-<br>diacylglycerol kinase<br>[Rattus | 5.5     |
| 921                                   | U72396         | Lycopersicon<br>esculentum class II<br>small heat shock<br>protein Le-HSP17.6<br>mRNA, complete cds | 0.006   | 2196567   | (D88588) lipoprotein<br>[Escherichia coli]  | 4.3     |
| 922                                   | AF074387       | Sambucus nigra<br>hevein-like protein<br>mRNA, complete cds   | 0.006   | 2113798   | (Z83259) AmphiBrf38<br>[Branchiostoma<br>floridae]  | 4.3     |
| 923                                   | AB012106       | Brassica rapa mRNA<br>for SRK45, complete<br>cds  | 0.006   | 1388166   | (U58282) Bowel<br>[Drosophila<br>melanogaster]  | 4.3     |
| 924                                   | AF074386       | Sambucus nigra<br>hevein-like protein<br>mRNA, complete cds   | 0.006   | 2496785   | HYPOTHETICAL<br>20.1 KD PROTEIN<br>Y4YS   | 4.2     |
| 925                                   | AF012899       | Sambucus nigra<br>ribosome inactivating<br>protein precursor<br>mRNA, complete cds                  | 0.006   | 416592  | A-AGGLUTININ<br>ATTACHMENT<br>SUBUNIT<br>PRECURSOR<br>>gi 101170 pir  A4125<br>8 a-agglutinin core<br>protein AGA1 - yeast<br>(Saccharomyces<br>cerevisiae)                               | 2.7     |
| 926                                   | AF064029       | Helianthus tuberosus<br>lectin 1 mRNA,<br>complete cds  | 0.006   | 416592  | A-AGGLUTININ<br>ATTACHMENT<br>SUBUNIT<br>PRECURSOR<br>>gi 101170 pir  A4125<br>8 a-agglutinin core<br>protein AGA1 - yeast<br>(Saccharomyces<br>cerevisiae)                               | 2.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 927    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 0.006   | 3258584  | (U41263) The 3' UTR of this gene overlaps the 3' UTR of T19D12.6(confirmed by EST hits) [Caenorhabditis elegans] | 2.0     |
| 928    | U33949                                | Human Down Syndrome region of chromosome 21, genomic sequence, clone A12H1-1A6.      | 0.006   | 3850997  | (AF067150) beta-hydroxyacyl-ACP dehydratase precursor  | 1.9     |
| 1175   | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1176   | Y09232                                | H.sapiens fertilin alpha pseudogene  | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1177   | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1178   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1179   | AF072847                              | Homo sapiens putative swelling-activated chloride channel (CLNSIA) gene, intron 6    | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1180   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1181   | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete ads            | 2e-04   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1182   | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds                  | 2e-04   | 1213557  | (U50199) coded for by C. elegans cDNA yk89e9.5; coded for by C. elegans cDNA cm7g5; coded for by C. elegans cDNA cm14b9; coded for by C. elegans cDNA yk52g5.5; coded for by C. elegans cDNA yk76e5.5; coded for by C. elegans cDNA yk131f11.5; c... | 8.4     |
| 1183   | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 2e-04   | 729008   | EPITHELIAL DISCOIDIN DOMAIN RECEPTOR I PRECURSOR (TYROSINE-PROTEIN KINASE CAK) (CELL ADHESION KINASE) (TYROSINE KINASE DDR) (DISCOIDIN RECEPTOR TYROSINE KINASE) (TRK E) (PROTEIN-TYROSINE KINASE RTK 6) sapiens]                                    | 8.3     |
| 1184   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                             | 2e-04   | 2507582  | HYPOTHETICAL 138.1 KD PROTEIN IN MOLR-BGLX INTERGENIC REGION >gi 1788436 (AE000300) putative regulator [Escherichia coli]  | 7.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 1185   | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds                                | 2e-04   | 1085500  | collagen alpha 1 (IX) chain - mouse musculus] >gi 744962 prf  2015346A collagen:SUBUNIT= alpha 1:ISO TYPE=IX [Mus musculus]   | 7.8     |
| 1186   | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 2e-04   | 2623967  | (Y13942) GTN Reductase [Agrobacterium radiobacter]  | 7.4     |
| 1187   | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 2e-04   | 2497316  | ADVANCED GLYCOSYLATION END PRODUCT-SPECIFIC RECEPTOR PRECURSOR (RECEPTOR FOR ADVANCED GLYCOSYLATION END PRODUCTS) products receptor precursor - bovine>gi 163651 (M91212) receptor for advanced glycosylation end products [Bos taurus] | 5.3     |
| 1188   | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 2e-04   | 1001710  | (D64004) hypothetical protein   | 3.5     |
| 1189   | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 2e-04   | 3258584  | (U41263) The 3' UTR of this gene overlaps the 3' UTR of T19D12.6 (confirmed by EST hits) [Caenorhabditis elegans]   | 2.1     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1190                                  | AF027173       | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds                  | 2e-04   | 2736338   | (AF038623_ contains similarity to RNA recognition motifs                           | 0.89    |
| 1191                                  | U72396         | Lycopersicon esculentum class II small heat shock protein Le-HSP17.6 mRNA, complete cds               | 2e-04   | 2196567   | (D88588) lipoprotein [Escherichia coli]  | 0.69    |
| 1192                                  | AF090115       | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 2e-04   | 3319874   | (AJ006096) F-spondin [Branchiostoma floridae]                                      | 5e-04   |
| 1193                                  | L26049         | Chlamydomonas reinhardtii dynein heavy chain alpha (ODA11) gene, exons 2-15, and partial cds.         | 2e-04   | 3876775   | (Z81077) predicted using Genefinder; Similarity to Yeast protein 8248 (TR:G587531) | 2e-09   |
| 1194                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds  | 1e-04   | <NONE>  | <NONE>   | <NONE>  |
| 1195                                  | AF064029       | Helianthus tuberosus lectin 1 mRNA, complete cds  | 1e-04   | <NONE>  | <NONE>   | <NONE>  |
| 1196                                  | L34219         | Homo sapiens retinaldehyde-binding protein (CRALBP) gene, complete cds.                               | 1e-04   | <NONE>  | <NONE>   | <NONE>  |
| 1197                                  | X51890         | Rhesus monkey interleukin-3 gene  | 1e-04   | <NONE>  | <NONE>   | <NONE>  |
| 1198                                  | AE001421       | Plasmodium falciparum chromosome 2, section 58 of 73 of the complete sequence                         | 1e-04   | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 1199   | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP 17.4) mRNA, complete cds | 1e-04   | <NONE>   | <NONE>  | <NONE>  |
| 1200   | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds                   | 1e-04   | 2576287  | (Y15086) HepC protein [Cylindrotheca fusiformis]  | 4.7     |
| 1201   | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme   | 1e-04   | 3395673  | (AB016623) RWC-3 [Oryza sativa]   | 0.14    |
| 1202   | AF038035                              | Homo sapiens BRCA1-associated RING domain protein (BARD1) gene, exons 2 and 3                          | 9e-05   | <NONE>   | <NONE>  | <NONE>  |
| 1203   | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme   | 9e-05   | <NONE>   | <NONE>  | <NONE>  |
| 1204   | AB012106                              | Brassica rapa mRNA for SRK45, complete cds   | 9e-05   | <NONE>   | <NONE>  | <NONE>  |
| 1205   | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds   | 9e-05   | <NONE>   | <NONE>  | <NONE>  |
| 1206   | AF034099                              | Laccaria bicolor glyoxal malate synthase protein mRNA, complete cds                                    | 9e-05   | 1351553  | HYPOTHETICAL LIPOPROTEIN MG348 PRECURSOR >gi 1361668 pir E6423 8 hypothetical protein MG348 - Mycoplasma genitalium (SGC3) gi 3844931 | 8.8     |
| 1207   | D50006                                | Human DNA for alpha-platelet-derived growth factor receptor, exon 6-10                                 | 9e-05   | 3063639  | (AF056494) NADH dehydrogenase subunit 5 [Panorpa japonica]  | 5.1     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1208                                  | U50423         | Human Down Syndrome region of chromosome 21, clone A41B8-1B7.                       | 9e-05   | 124273  | INHIBIN ALPHA CHAIN PRECURSOR bovine >gi 163195 (M13273) inhibin A subunit [Bos taurus] | 3.0     |
| 1209                                  | AJ005813       | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                            | 9e-05   | 4007782   | (X72850) 2.4-dihydroxybenzoate monooxygenase [Sphingomonas sp.]                         | 2.3     |
| 1210                                  | AC005276       | Homo sapiens clone fragment UWGC:gap3 from 7q31.3, complete sequence [Homo sapiens] | 9e-05   | 1492075   | (U60315) MC132L [Molluscum contagiosum virus subtype 1]                                 | 1.0     |
| 1211                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds  | 9e-05   | 2887423   | (AB007884) KIAA0424 [Homo sapiens]  | 2e-10   |
| 1212                                  | X77772         | C.fuscus gamma-M2-1 crystallin mRNA.  | 9e-05   | 2072425   | (U83115) non-lens beta gamma-crystallin like protein [Homo sapiens]                     | 7e-25   |
| 1213                                  | AB012106       | Brassica rapa mRNA for SRK45, complete cds  | 8e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1214                                  | L06178         | Apis mellifera ligustica complete mitochondrial genome                              | 8e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1215                                  | AB012106       | Brassica rapa mRNA for SRK45, complete cds  | 8e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1216                                  | U76524         | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds           | 8e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1217                                  | L06178         | Apis mellifera ligustica complete mitochondrial genome                              | 8e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1218                                  | AB012106       | Brassica rapa mRNA for SRK45, complete cds  | 8e-05   | <NONE>  | <NONE>  | <NONE>  |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1219                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds   | 8e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1220                                  | AB012106       | Brassica rapa mRNA for SRK45, complete cds   | 8e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1221                                  | AB012106       | Brassica rapa mRNA for SRK45, complete cds   | 8e-05   | 1722841   | WNT-11 PROTEIN PRECURSOR (XWNT-11) clawed frog >gi 439108 (L23542) maternal protein  | 9.9     |
| 1222                                  | AF027174       | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 8e-05   | 1205991   | (U35637) nebulin [Homo sapiens]  | 9.6     |
| 1223                                  | AF024605       | Homo sapiens serine protease-like protease Sequence 2 from patent US 5736377         | 8e-05   | 3242783   | (AF055354) respiratory burst oxidase protein B   | 8.6     |
| 1224                                  | Y13148         | Rattus norvegicus mRNA of PAG608 gene  | 8e-05   | 2314243   | (AE000616) alpha-ketoglutarate permease (kgtP)   | 8.1     |
| 1225                                  | AJ005813       | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 8e-05   | 1170586   | RAS GTPASE-ACTIVATING-LIKE PROTEIN IQGAP1 (P195) (KIAA0051) >gi 627594 pir  A5485 4 Ras GTPase activating-related protein - human sapiens] >gi 536844 (L33075) ras GTPase-activating-like protein [Homo sapiens] | 7.8     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1226                                  | AF027173       | Arabidopsis thaliana<br>cellulose synthase<br>catalytic subunit (Ath-<br>A) mRNA, complete<br>cds | 8e-05   | 464239  | NADH-<br>UBIQUINONE<br>OXIDOREDUCTASE<br>CHAIN<br>4>gi 1085185 pir  S52<br>968 NADH<br>dehydrogenase chain 4<br>- honeybee<br>mitochondrion<br>(SGC4) >gi 552446<br>(L06178) NADH<br>dehydrogenase subunit<br>4 [Apis mellifera<br>ligustica] | 3.5     |
| 1227                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds  | 8e-05   | 544353  | F-SPONDIN<br>PRECURSOR  | 3.5     |
| 1228                                  | U76524         | Sambucus nigra<br>ribosome inactivating<br>protein precursor<br>mRNA, complete cds                | 8e-05   | 483243  | apolipoprotein B-100 -<br>chicken (fragment)  | 3.4     |
| 1229                                  | AF093268       | Rattus norvegicus<br>homer-1c mRNA,<br>complete cds   | 8e-05   | 91207   | proline-rich protein -<br>mouse (fragment)<br>musculus]   | 22      |
| 1230                                  | AF027173       | Arabidopsis thaliana<br>cellulose synthase<br>catalytic subunit (Ath-<br>A) mRNA, complete<br>cds | 8e-05   | 2499181   | ZONADHESIN<br>PRECURSOR<br>>gi 1066466  | 2.2     |
| 1231                                  | AF027173       | Arabidopsis thaliana<br>cellulose synthase<br>catalytic subunit (Ath-<br>A) mRNA, complete<br>cds | 8e-05   | 2499181   | ZONADHESIN<br>PRECURSOR<br>>gi 106646   | 1.9     |
| 1232                                  | AB012106       | Brassica rapa mRNA<br>for SRK45, complete<br>cds  | 8e-05   | 2833647   | (AF027972)<br>flagelliform silk<br>protein [Nephila<br>clavipes]  | 1.6     |
| 1233                                  | AF093268       | Rattus norvegicus<br>homer-1c mRNA,<br>complete cds   | 8e-05   | 1163063   | (Z49821) MYO2<br>[Saccharomyces<br>cerevisiae]  | 0.90    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1234   | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds                  | 8e-05   | 1653488  | (D90914) hypothetical protein  | 0.30    |
| 1235   | M26510                                | Chicken nonmuscle myosin heavy chain (MHC) gene, complete cds.  | 8e-05   | 112159   | plectin - rate   | 0.003   |
| 1236   | U56402                                | Human chromatin structural protein homolog  | 8e-05   | 2088823  | (AF003384) weak similarity to the peptidase family A2  | 1e-13   |
| 1237   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 8e-05   | 437181   | (U02289) GTPase-activating protein [Caenorhabditis elegans]  | 2e-17   |
| 1238   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 8e-05   | 465983   | HYPOTHETICAL 80.8 KD PROTEIN ZC21.4 IN CHROMOSOME III  | 8e-27   |
| 1239   | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 7e-05   | <NONE>   | <NONE>   | <NONE>  |
| 1240   | U83656                                | Rattus norvegicus NF-KB gene, promotor region   | 7e-05   | 3880858  | (AL031633) predicted using Genefinder; cDNA EST yk304f12.5 comes from this gene [Caenorhabditis elegans] | 9.3     |
| 1241   | AF074387                              | Sambucus nigra hevein-like protein mRNA, complete cds   | 7e-05   | 3080538  | (A1022600) hypothetical protein  | 9.2     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1242                                  | X89398         | H.sapiens ung gene for uracil DNA-glycosylase                             | 7e-05   | 549700  | HYPOTHETICAL 23.7 KD PROTEIN IN MDH1-VMA5 INTERGENIC REGION<br>>gi 539182 pir  S3790 8 hypothetical protein YKL083w - yeast (Saccharomyces cerevisiae) >gi 486120 (Z28082) ORF YKL083w | 1.8     |
| 1243                                  | M83753         | Bovine follicle stimulating hormone-beta subunit gene, complete cds.      | 7e-05   | 2398621   | (AJ000342) DMBT1 protein, 5.8 kb transcript [Homo sapiens]   | 1.8     |
| 1244                                  | M80829         | Rat troponin T cardiac isoform gene, complete cds                         | 5e-05   | 854065  | (X83413) U88 [Human herpes virus 6]  | 2e-08   |
| 1245                                  | AF074387       | Sambucus nigra hevein-like protein mRNA, complete cds                     | 4e-05   | 120240  | FLAGELLIN B2 PRECURSOR Methanococcus voltae >gi 150063 (M72148) flagellin  | 5.2     |
| 1246                                  | AF012899       | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 3e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1247                                  | U76524         | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 3e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1248                                  | AF074386       | Sambucus nigra hevein-like protein mRNA, complete cds                     | 3e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1249                                  | AF093268       | Rattus norvegicus homer-1c mRNA, complete cds                             | 3e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1250                                  | AB012106       | Brassica rapa mRNA for SRK45, complete cds                                | 3e-05   | 2773226   | (AF039713) Similar to protein kinase [Caenorhabditis elegans]  | 6.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1251   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 3e-05   | 2072961  | (U93568)putative p150 [Homo sapiens]   | 5.6     |
| 1252   | U72396                                | Lycopersicon esculentum class II small heat shock protein Le-HSP17.6 mRNA, complete cds | 3e-05   | 121855   | EXOGLUCANASE II PRECURSOR cellulose 1,4-beta cellobiosidase (EC 3.2.1.91) II precursor - fungus (Trichoderma reesei) 1,4-beta-cellobiosidase (EC 3.2.1.91) II - fungus cellobiohydrolase II [Trichoderma reesei] | 4.6     |
| 1253   | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds               | 3e-05   | 3880516  | (AL021572) similar to CTP SYNTHASE (EC 6.3.4.2) (UTP--AMMONIA LIGASE) (CTP SYNTHETASE)   | 3.3     |
| 1254   | M88299                                | Mouse brain-1 POU-domain protein, complete cds.   | 3e-05   | 1947048  | (U66102) intimin [Escherichia coli]  | 3.0     |
| 1255   | U95098                                | Xenopus laevis mitotic phosphoprotein 44 mRNA, partial cds                              | 3e-05   | 3122872  | CELL-CYCLE NUCLEAR AUTOANTIGEN SG2NA (S/G2 NUCLEAR ANTIGEN) >gi 1082650 pir  JC25 22 nuclear autoantigen - human >gi 805095 (U17989) GS2NA   | 2.8     |
| 1256   | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds               | 3e-05   | 1352145  | CYTOCHROME C OXIDASE POLYPEPTIDE I chain I - Thermus aquaticus >gi 155083 (M84341) cytochrome c oxidase subunits precursor [Thermus thermophilus]  | 2.6     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1257                                  | U72396         | Lycopersicon<br>esculentum class II<br>small heat shock<br>protein Le-HSP17.6<br>mRNA, complete cds | 3e-05   | 2811015   | SEGMENTATION<br>POLARITY<br>PROTEIN<br>ENGRAILED<br>>gi 2076747 (U42429)<br>engrailed [Anopheles<br>gambiae] >gi 2148918<br>(U42214) engrailed<br>[Anopheles gambiae]     | 2.0     |
| 1258                                  | AF027174       | Arabidopsis thaliana<br>cellulose synthase<br>catalytic subunit (Ath-<br>B) mRNA, complete<br>cds   | 3e-05   | 1657752   | (U62325) FE65-like<br>protein [Homo<br>sapiens]   | 1.7     |
| 1259                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds  | 3e-05   | 2072961   | (U93568) putative<br>p150 [Homo sapiens]  | 1.5     |
| 1260                                  | U76523         | Sambucus nigra lectin<br>precursor mRNA,<br>complete cds  | 3e-05   | 1352145   | CYTOCHROME C<br>OXIDASE<br>POLYPEPTIDE I<br>chain I - Thermus<br>aquaticus >gi 155083<br>(M84341) cytochrome<br>c oxidase subunits<br>precursor [Thermus<br>thermophilus] | 1.1     |
| 1261                                  | X91890         | H.sapiens regulatory<br>region of HOXA7<br>gene   | 3e-05   | 111013  | Sxr (Bkm-homolog)<br>sex-determining<br>region protein - mouse  | 1.0     |
| 1262                                  | L36936         | Homo sapiens metase<br>gene, partial cds.   | 3e-05   | 1944352   | (D84239) IgG Fc<br>binding protein<br>[Homo sapiens]  | 0.99    |
| 1263                                  | AB012105       | Brassica rapa mRNA<br>for SLG45, complete<br>cds  | 3e-05   | 417782  | SMP2 PROTEIN<br>>gi 320853 pir  S3091<br>1 SMP2 protein -<br>yeast (Saccharomyces<br>cerevisiae) gene<br>[Saccharomyces<br>cerevisiae]                                    | 0.89    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1264   | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                             | 3e-05   | 1708501  | INTEGRIN ALPHA CHAIN-LIKE PROTEIN alpha Intlp [Candida albicans]   | 0.39    |
| 1265   | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 3e-05   | 1587031  | cis-Golgi matrix protein GM130 [Rattus norvegicus]   | 0.20    |
| 1266   | Z81014                                | Human DNA sequence from cosmid U65A4, between markers DXS366 and DXS87 on chromosome X *              | 3e-05   | 2072964  | (U93569) putative p150 [Homo sapiens]  | 0.049   |
| 1267   | Z96668                                | H.sapiens telomeric DNA sequence, clone 7PTEL001, read 7PTELOO001.seq                                 | 3e-05   | 542429   | glycosylated and myristilated smaller surface antigen - Plasmodium falciparum >gi 836640 (X76298) glycosylated and myristilated smaller surface antigen gallus] >gi 1092178 prf  2023 165B surface antigen | 0.029   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 1268   | AB012105                              | Brassica rapa mRNA for SLG45, complete cds   | 3e-05   | 3879121  | (Z70310) predicted using Genefinder; Similarity to Muse ankyrin (PIR Acc. No. S37771); cDNA EST EMBL:T01923 comes from this gene; cDNA EST EMBL:D32335 comes from this gene; cDNA EST EMBL:D32723 comes from this gene; cDNA ES... Genefinder; Similarity to Mouse ankyrin (PIR Acc. No. S37771); cDNA EST EMBL:T01923 comes from this gene; cDNA EST EMBL:D32335 comes from this gene; cDNA EST EMBL:D32723 comes from this gene; cDNA ES... | 2e-13   |
| 1269   | AF074385                              | Sambucus nigra hevein-like protein mRNA, complete cds                                | 3e-05   | 2497677  | ZYXIN (ZYXIN 2) sapiens] >gi 1545954 gnl PID e 223417 (X95735) zyxin  | 2e-23   |
| 1270   | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 1e-05   | <NONE>   | <NONE>  | <NONE>  |
| 1271   | X16318                                | Canine mRNA for signal recognition particle 54k protein                              | 1e-05   | 3122612  | PITUITARY HOMEODOMAIN 3 (HOMEODOMAIN PROTEIN PITX3) >gi 2645427 (AF005772) homeobox protein Pitx3 [Mus musculus]  | 4.4     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1272   | AB012105                              | Brassica rapa mRNA for SLG45, complete cds  | 1e-05   | 1652458  | (D90905) DNA mismatch repair protein MutL [Synechocystis sp.]  | 0.62    |
| 1273   | U57843                                | Human phosphatidylinositol 3-kinase delta catalytic subunit mRNA, complete cds                              | 1e-05   | 475909   | (X67098) ORF1A [Homo sapiens]  | 0.22    |
| 1274   | Z96569                                | H.sapiens telomeric DNA sequence, clone 2QTEL054, read 2QTELOO054, seq                                      | 1e-05   | 2137043  | unknown protein - rabbit (fragment) cuniculus]   | 0.005   |
| 1275   | AE000810                              | Methanobacterium thermoautotrophicum from bases 172512 to 182957 (section 16 of 148) of the complete genome | 1e-05   | 3877579  | (Z02271) Similarity to Mouse kinensin-like protein KIF4 (SW:P33174); cDNA EST EMBL:D27320 comes from this gene; cDNA EST EMBL:D27322 comes from this gene; cDNA EST EMBL:D27321 comes from this gene; cDNA EST EMBL:D35764 comes... Mouse kinensin-like protein KIF4 (SW:P33174); cDNA EST EMBL:D27320 comes from this gene; cDNA EST EMBL:D27322 comes from this gene; cDNA EST EMBL:D27321 comes from this gene; cDNA EST EMBL:D35764 comes... | 6e-27   |
| 1276   | AB012113                              | Homo sapiens gene for CC chemokine PARC precursor, complete cds   | 9e-06   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1277   | AC005830                              | Homo sapiens Xp22-154-155 BAC GSHB-524I1 (Genome Systems Human BAC Library), complete sequence [Homo sapiens] | 9e-06   | <NONE>   | <NONE>  | <NONE>  |
| 1278   | D86245                                | Human MHC (HLA) DRB intron 1 DNA, partial sequence  | 9e-06   | 1051253  | (U37531) mucin apoprotein [Mus musculus]  | 1.3     |
| 1279   | D79998                                | Human mRNA for KIAA0176 gene, partial cds   | 9e-06   | 2833253  | HYPOTHETICAL PROTEIN KIAA0176 sapiens]  | 4e-06   |
| 1280   | U10246                                | Toxoplasma gondii RH uracil phosphoribosyl transferase gene, complete cds.                                    | 9e-06   | 3876090  | (Z69635) Similarity to Yeast uridine kinase (SW:URK1_YEAST); cDNA EST EMBL:Z14695 comes from this gene; cDNA EST CEMSE17F comes from this gene; cDNA EST EMBL:D67355 comes from this gene; cDNA EST yk209h1.5 comes from this ge... | 7e-33   |
| 1281   | U10646                                | Toxoplasma gondii RH uracil phosphoribosyl transferase gene, complete cds.                                    | 9e-06   | 3876090  | (Z69635) Similarity to Yeast uridine kinase (SW:URK1_YEAST); cDNA EST EMBL:Z14695 comes from this gene; cDNA EST CEMSE17F comes from this gene; cDNA EST EMBL:D67355 comes from this gene; cDNA EST yk209h1.5 comes from this ge... | 7e-34   |
| 1282   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                                     | 8e-06   | <NONE>   | <NONE>  | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1283                                  | AF012899       | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                          | 8e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1284                                  | U66240         | Human Rh blood group C antigen (RHCE) gene, exon 2, partial cds                                    | 8e-06   | 1707155   | (U80837) F07E5.6 gene product [Caenorhabditis elegans] | 9.6     |
| 1285                                  | AF012899       | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                          | 7e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1286                                  | M29930         | Human insulin receptor (allele 2) gene, exons 14, 15, 16 and 17.                                   | 4e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1287                                  | L42103         | Homo sapiens (subclone 5_d3 from P1 H25) DNA sequence.   | 3e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1288                                  | AF012244       | Mus musculus cerberus-like (Cer-1) gene, complete cds  | 3e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1289                                  | Z69366         | Human DNA sequence from cosmid L96F8, Huntington's Disease Region, chromosome 4p16.3 contains EST. | 3e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1290                                  | Z69366         | Human DNA sequence from cosmid L96F8. Huntington's Disease Region, chromosome 4p16.3 contains EST. | 3e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1291                                  | X85232         | H.sapiens chromosome 3 sequences   | 3e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1292                                  | M32674         | Human platelet glycoprotein IIIa. exons 7, 8 and 9.  | 3e-06   | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 1293   | D16879                                | Human HepG2 3' region cDNA, clone hmd2a01   | 3e-06   | 998296   | (U33484) ependymin [Hemiodus sp.]   | 5.6     |
| 1294   | U18614                                | Lagothrix lagotricha interphotoreceptor retinoid-binding protein (IRBP) gene, intron 1, complete sequence | 3e-06   | 1613846  | (U71440) polyprotein [Rice tungro spherical virus]  | 5.0     |
| 1295   | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds     | 3e-06   | 1477646  | (U53204) plectin [Homo sapiens]<br>>gi 1477651 (U63610) plectin [Homo sapiens]  | 4.0     |
| 1296   | AF016898                              | Homo sapiens B-ATF gene, complete cds   | 3e-06   | 1085177  | reverse transcriptase - fruit fly reverse transcriptase [Drosophila yakuba]   | 3.0     |
| 1297   | AB018490                              | Homo sapiens DNA, trinucleotide repeats region  | 3e-06   | 3876572  | (Z81522) predicted using Genefinder; similar to RNA recognition motif. (aka RRM, RBD, or RNP domain) [Caenorhabditis elegans] | 3.0     |
| 1298   | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds                      | 3e-06   | 4240137  | (AB020631) KIAA0824 protein [Homo sapiens]  | 2.7     |
| 1299   | M37929                                | Homo sapiens adenosine monophosphate deaminase 1 (AMPD1) gene, exons 11-12.                               | 3e-06   | 1653775  | (D90916) thiol:disulfide interchange protein DsbD [Synechocystis sp.]   | 1.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1300   | M37929                                | Homo sapiens adenosine monophosphate deaminase 1 (AMPD1) gene, exons 11-12.  | 3e-06   | 1653775  | (D90916) thiol:disulfide interchange protein DsbD [Synechocystis sp.]      | 1.7     |
| 1301   | U60496                                | Glycine max actin (Soy86) gene, partial cds  | 3e-06   | 173078   | ACTIN-LIKE PROTEIN ARP5 Ynl2430[ [Saccharomyces cerevisiae]                | 2e-05   |
| 1302   | X14363                                | Yersinia pseudotuberculosis rplC, rplD, rplW, rplB and rpsS genes for ribosomal proteins L3, L4, L23, L2 and S19     | 3e-06   | 585879   | 50S RIBOSOMAL PROTEIN L2 maritima >gi 437926 (Z21677) ribosomal protein L2 | 2e-12   |
| 1303   | Z34969                                | H.sapiens DNA for microsatellite polymorphism  | 2e-06   | <NONE>   | <NONE>   | <NONE>  |
| 1304   | X64707                                | H.sapiens BBC1 mRNA  | 1e-06   | <NONE>   | <NONE>   | <NONE>  |
| 1305   | AC005830                              | Homo sapiensXp22-154-155 BAC GSHB-52411 (Genome Systems Human BAC Library), complete sequence [Homo sapiens]         | 1e-06   | <NONE>   | <NONE>   | <NONE>  |
| 1306   | J04058                                | Human electron transfer flavoprotein alpha-subunit mRNA, complete cds.   | 1e-06   | <NONE>   | <NONE>   | <NONE>  |
| 1307   | L25647                                | Homo sapiens fibroblast growth factor receptor gene (located in the central MHC) signal peptide and consecutive exon | 1e-06   | 1586734  | mxoQ gene [Methylobacterium organophilum]                                  | 5.4     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 1308   | L26261                                | Human MHC class III HLA-RP1 gene.   | 1e-06   | 1684985  | (U20633) NADH dehydrogenase subunit [Neuwiedia veratrifolia]  | 1.8     |
| 1309   | AF002283                              | Mus musculus alpha-actinin-2 associated LIM protein mRNA, alternatively spliced product, complete cds             | 1e-06   | 2996196  | (AF053367) carboxyl terminal LIM domain protein [Mus musculus]  | 4e-17   |
| 1310   | M10935                                | Human haptoglobin gene (alpha-2 allele), complete cds and haptoglobin-related gene, exon 1 and three Alu repeats. | 6e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1311   | AC002251                              | Homo sapiens (subclone 1_g6 from BAC H76) DNA sequence  | 4e-07   | 2144491  | coagulation factor Xa (EC3.4.21.6) precursor norvegicus]  | 4.2     |
| 1312   | AF047717                              | Streptomyces chrysomallus actinomycin synthetase II (acmB) gene, complete cds                                     | 4e-07   | 699196   | (U15181) 4-coumarate-coA ligase [Mycobacterium leprae]  | 1e-06   |
| 1313   | U14417                                | Human Ral guanine nucleotide dissociation stimulator mRNA, partial cds.   | 4e-07   | 544402   | GUANINE NUCLEOTIDE DISSOCIATION STIMULATOR RALGDS FORM A (RALGEF)<br>>gi 321257 pir  S2841 5 guanine nucleotide dissociation stimulator ralGDS - mouse<br>>gi 193573 (L07924) guanine nucleotide dissociation stimulator [Mus musculus] | 8e-08   |
| 1314   | Z79027                                | H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA20G8  | 3e-07   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1315   | U67167                                | Homo sapiens intestinal mucin (MUC2) gene, promoter region and partial cds                                    | 3e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1316   | AF086256                              | Homo sapiens full length insert cDNA clone ZD41C11  | 3e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1317   | U67228                                | Human clone HS4.61 Alu-Ya5 sequence   | 3e-07   | 1938437  | (U97003) contains similarity to C4-type zinc fingers and a ligand-binding domain of nuclear hormone receptors | 2.3     |
| 1318   | U94346                                | Human calpain-like protease (htra-3) mRNA, complete cds   | 3e-07   | 2911858  | (AF047659) No definition line found [Caenorhabditis elegans]  | 0.39    |
| 1319   | Y15724                                | Homo sapiens SERCA3 gene, exons 1-7 (and joined CDS)  | 1e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1320   | X13596                                | Bean DNA for glycine-rich cell wall protein GRP 1.8   | 1e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1321   | M83094                                | Homo sapiens cytosolic selenium-dependent glutathione peroxidase gene, complete cds, and rhoh12 gene, 3' end. | 1e-07   | 1326385  | (U58751) C07G1.7 gene product [Caenorhabditis elegans]  | 8.0     |
| 1322   | Z55905                                | H.sapiens CpG DNA, clone 71f4, forward read cpg71f4.ftla.   | 1e-07   | 1076802  | extensin-like protein - maize >gi 600118 mays]  | 0.61    |
| 1323   | X03541                                | Human mRNA of trk oncogene > :: gb I96186 I96186 Sequence 23 from patent US 5734039                           | 1e-07   | 325465   | (M74509) [Human endogenous retrovirus type C oncovirus sequence.], gene product [Homo sapiens]                | 3e-04   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1324                                  | AF027766       | Canis familiaris Y-linked zinc finger protein  | 1e-07   | 220643  | (D10628) zinc finger protein [Mus musculus]   | 7e-08   |
| 1325                                  | D13613         | Bovine mRNA for rabphilin-3A, complete cds > :: dbj E07809 E07809 cDNA encoding rabphilin-3A | 1e-07   | 2822161   | (AC004082) rab3 effector-like; 35% Similarity to AF007836 (PID:g2317778) [Homo sapiens]   | 6e-11   |
| 1326                                  | X57110         | Human mRNA for c-cbl proto-oncogene  | 1e-07   | 323270  | (J04169) gag-onc fusion protein [Cas NS1 retrovirus]  | 3e-14   |
| 1327                                  | X57110         | Human mRNA for c-cbl proto-oncogene  | 1e-07   | 115855  | PROTO-ONCOGENE C-CBL human >gi 29731 (X57110) c-cbl protein [Homo sapiens]  | 4e-19   |
| 1328                                  | AC001178       | Homo sapiens (subclone 2_g12 from BAC H94) DNA sequence                                      | 4e-08   | <NONE>  | <NONE>  | <NONE>  |
| 1329                                  | U11866         | Human interleukin-8 receptor type B (IL8RB) gene, promoter and exons 1-6                     | 4e-08   | <NONE>  | <NONE>  | <NONE>  |
| 1330                                  | AC001225       | Homo sapiens (subclone 2_e6 from BAC H94) DNA sequence                                       | 4e-08   | 478184  | histone H1 II-1 (clone L95) - midge   | 6.5     |
| 1331                                  | M73837         | Human modulator recognition factor 2 (MRF-2) mRNA, complete cds.                             | 4e-08   | 141448  | HYPOTHETICAL 32.6 KD PROTEIN IN TRANSPOSON TN4556 >gi 80758 pir  JQ0428 hypothetical 32.6K protein - Streptomyces fradiae transposon Tn4556 | 4.7     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1332   | AC006164                              | Homo sapiens clone UWGC:y28gap from 6p21, complete sequence [Homo sapiens]                         | 4e-08   | 2580578  | (AF000996) ubiquitous TPR motif, Y isoform [Homo sapiens]  | 1.2     |
| 1333   | X01060                                | Human mRNA for transferrin receptor  | 4e-08   | 135514   | T-CELL RECEPTOR BETA CHAIN PRECURSOR precursor (ANA 11) rabbit   | 0.61    |
| 1334   | Y10697                                | H.sapiens INE2 mRNA  | 4e-08   | 127909   | INSULIN RECEPTOR - RELATED PROTEIN PRECURSOR (IRR) IR-RELATED RECEPTOR) >gi 186555 sapiens]  | 0.14    |
| 1335   | U60416                                | Rattus norvegicus myr 6 myosin heavy chain mRNA, complete cds                                      | 4e-08   | 102189   | myosin I, high molecular weight - Acanthamoeba sp  | 3e-08   |
| 1336   | U23804                                | Drosophila melanogaster putative GTP-binding regulatory protein beta chain (GPB) mRNA partial cds. | 4e-08   | 2494916  | HYPOTHETICAL 55.2 KD TRP-ASP REPEATS CONTAINING PROTEIN T10F2.4 IN CHROMOSOME III protein; similar to G-Beta repeat region (Trp-Asp domains) of guanine nucleotide binding protein | 1e-28   |
| 1337   | AE000213                              | Escherichia coli K-12 MG1655 section 103 of 400 of the complete genome                             | 4e-08   | 3294172  | (AL022325) tF27C3.1.1 (protein similar to C. elegans protein B0035.16) (isoform 1) [Homo sapiens]  | 2e-67   |
| 1338   | D89821                                | Mus musculus mRNA for RhoM, complete cds   | 2e-08   | 3024539  | RHO-RELATED GTP-BINDING PROTEIN RHOD (RHO-RELATED PROTEIN HP1) (RHOHP1) sapiens]   | 1e-04   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1339   | U74382                                | Human telomeric repeat DNA-binding protein (PIN2) mRNA, complete cds | 1e-08   | <NONE>   | <NONE>   | <NONE>  |
| 1340   | L35657                                | Homo sapiens (subclone H8 5_a10 from P1 35 H5 C8) DNA sequence.      | 1e-08   | <NONE>   | <NONE>   | <NONE>  |
| 1341   | L21936                                | Human succinate dehydrogenase flavoprotein subunit                   | 1e-08   | 3201678  | (AF060886) adenine phosphoribosyltransferase [Leishmania tarentolae]   | 4.0     |
| 1342   | AB009777                              | Homo sapiens gene for osteonidogen, promoter region                  | 1e-08   | 479388   | tritin - wheat >gi 391929 gnl PID d1003454   | 2.2     |
| 1343   | M58600                                | Human heparin cofactor II (HCF2) gene, exons 1 through 5.            | 1e-08   | 1730173  | GLUCOSE-6-PHOSPHATE ISOMERASE, CYTOSOLIC 2 (GPI) (PHOSPHOGLUCOSE ISOMERASE) (PGI) isomerase [Clarkia concinna] | 1.9     |
| 1344   | M58600                                | Human heparin cofactor II (HCF2) gene, exons 1 through 5.            | 1e-08   | 1730173  | GLUCOSE-6-PHOSPHATE ISOMERASE, CYTOSOLIC 2 (GPI) (PHOSPHOGLUCOSE ISOMERASE) (PGI) isomerase [Clarkia concinna] | 1.7     |
| 1345   | AC000980                              | Homo sapiens (subclone 1_g2 from P1 H31) DNA sequence                | 1e-08   | 439877   | (L27428) reverse transcriptase [Homo sapiens]  | 1.1     |
| 1346   | U48734                                | Human non-muscle alpha-actinin mRNA, complete cds                    | 1e-08   | 168237   | (M76546) hydroxyproline-rich protein [Helianthus annuus]   | 0.19    |
| 1347   | M76724                                | Human leukocyte adhesion receptor alpha subunit                      | 1e-08   | 1177607  | (X92485) pval [Plasmodium vivax]   | 0.19    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 1348   | AF067959                              | Gallus gallus homeodomain protein HOXD-3 mRNA, complete cds                              | 1e-08   | 3165574  | (AF067942) No definition line found [Caenorhabditis elegans]   | 0.15    |
| 1349   | Z81014                                | Human DNA sequence from cosmid U65A4, between markers DXS366 and DXS87 on chromosome X * | 1e-08   | 2072964  | (U93569) putative p150 [Homo sapiens]  | 0.001   |
| 1350   | X57103                                | Human h-lys gene for lysozyme (upstream region)  | 7e-09   | <NONE>   | <NONE>   | <NONE>  |
| 1351   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                | 7e-09   | 231629   | BILE-SALT-ACTIVATED LIPASE PRECURSOR (ESTER LIPASE) (STEROL ESTERASE) (CHOLESTEROL ESTERASE) salt-activated lipase [Homo sapiens]  | 0.22    |
| 1352   | L34741                                | Aplysia californica prohormone convertase (PC2) mRNA, complete cds.                      | 5e-09   | 322054   | cytochrome-c oxidase (EC 1.9.3.1) chain II precursor - Synechocystis sp. (PCC 6803) >gi 581739 sp.]  | 5.0     |
| 1353   | AF052959                              | Homo sapiens type XV collagen (COL15A1) gene, exon 6                                     | 4e-09   | 131269   | PHOTOSYSTEM II P680 CHLOROPHYLL A APOPROTEIN (CP-47 PROTEIN) >gi 72708 pir QJLV6 A photosystem II chlorophyll a-binding protein psbB - liverwort (Marchantia polymorpha) chloroplast >gi 11700 | 1.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1354   | L15470                                | Streptomyces clavuligerus (NRRL 3585) clavulanic acid biosynthesis protein (cla) gene, complete cds and clavamate synthase 2 (cs2) gene, partial cds. | 4e-09   | 586028   | POSSIBLE AGMATINASE (AGMATINE UREOHYDROLASE) (AUH) (PROCLAVAMINIC ACID AMIDINO HYDROLASE) >gi 1361423 pir  S57669 Proclavaminic acid amidino hydrolase - Streptomyces clavuligerus >gi 295171 Proclavaminic acid amidino hydrolase [Streptomyces clavuligerus] >gi 1586122 prf  2203286B proclavaminic acid amidino hydrolase [Streptomyces clavuligerus] | 4e-13   |
| 1355   | AB002302                              | Human mRNA for KIAA0304 gene, complete cds  | 2e-09   | 131600   | GENERAL SECRETION PATHWAY PROTEIN L product [Klebsiella pneumoniae] >gi 149311 (M32613) pull  | 2.5     |
| 1356   | L34219                                | Homo sapiens retinaldehyde-binding protein (CRALBP) gene, complete cds.   | 1e-09   | <NONE>   | <NONE>  | <NONE>  |
| 1357   | AB002320<br>2                         | Human mRNA for KIAA0304 gene, complete cds  | 1e-09   | 2224549  | (AB002302) KIAA0304 [Homo sapiens]  | 5.0     |
| 1358   | D85731                                | Homo sapiens HSPA1L gene for Heat shock protein 70 testis variant, 5'UTR, partial sequence  | 1e-09   | 1389766  | (U58658) unknown [Homo sapiens]   | 1.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1359   | AF064483                              | Homo sapiens natural resistance-associated macrophage protein 2 (NRAMP2) gene, exon 17, alternatively spliced non-IRE for, complete cds | 8e-10   | 113671   | !!!! ALU CLASS F WARNING ENTRY !!!!  | 0.72    |
| 1360   | AF002283                              | Mus musculus alpha-actinin-2 associated LIM protein mRNA, alternatively spliced product, complete cds                                   | 6e-10   | 2996196  | (AF053367) carboxyl terminal LIM domain protein [Mus musculus]   | 4e-21   |
| 1361   | M26220                                | African green monkey origin of replication  | 5e-10   | 2143455  | gene DMR-N9 protein - mouse (fragment)   | 8.8     |
| 1362   | Z78006                                | H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA7F10  | 4e-10   | 2072977  | (U93574) putative p150 [Homo sapiens]  | 0.005   |
| 1363   | U82303                                | Homo sapiens unknown protein mRNA, partial cds  | 2e-10   | 1825711  | (U88183) similar to the immunoglobulin superfamily, most similar to neural cell adhesion proteins [Caenorhabditis elegans] | 0.031   |
| 1364   | AF079764                              | Drosophila melanogaster enhancer of polycomb  | 2e-10   | 3757890  | (AF079764) enhancer of polycomb [Drosophila melanogaster]  | 1e-10   |
| 1365   | L24123                                | Homo sapiens NRF1 protein (NRF1) mRNA.  | 2e-10   | 3004573  | (AC004520) similar to NFE2-related transcription factors; similar to I48694 (PID:g2137676) [Homo sapiens]                  | 4e-53   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1366   | M19454                                | Orangutan alpha-globin gene duplicate region.                        | 1e-10   | 464239   | NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4<br>>gi 1085185 pir  S529<br>68 NADH dehydrogenase chain 4 - honeybee mitochondrion (SGC4) >gi 552446 (L06178) NADH dehydrogenase subunit 4 [Apis mellifera ligustica] | 6.0     |
| 1367   | D87117                                | House mouse; Musculus domesticus brain mRNA for SAP102, complete cds | 6e-11   | 473912   | (L31961) phosphoprotein [Mus cookii]   | 2.2     |
| 1368   | AC001002                              | Homo sapiens (subclone 2_h9 from P1 H39) DNA sequence                | 5e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1369   | AC001002                              | Homo sapiens (subclone 2_h9 from P1 H39) DNA sequence                | 5e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1370   | AB007874                              | Homo sapiens KIA A0414 mRNA, partial cds                             | 5e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1371   | AC001002                              | Homo sapiens (subclone 2_h9 from P1 H39) DNA sequence                | 5e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1372   | AC001002                              | Homo sapiens (subclone 2_h9 from P1 H39) DNA sequence                | 5e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1373   | AC001002                              | Homo sapiens (subclone 2_h9 from P1 H39) DNA sequence                | 5e-11   | <NONE>   | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1374                                  | AC001002       | Homo sapiens (subclone 2_h9 from P1 H39) DNA sequence   | 5e-11   | <NONE>  | <NONE>   | <NONE>  |
| 1375                                  | Z21852         | H.sapiens mRNA for HERV-K long terminal repeat  | 5e-11   | 419481  | gag polyprotein - human endogenous virus S71   | 4.6     |
| 1376                                  | AB007928       | Homo sapiens mRNA for KIAA0459 protein, partial cds   | 5e-11   | 2947238   | (AF051782) diaphanous 1 [Homo sapiens]   | 2.8     |
| 1377                                  | D87117         | House mouse; Musculus domesticus brain mRNA for SAP102, complete cds                                | 5e-11   | 473912  | (L31961) phosphoprotein [Mus cookii]   | 1.8     |
| 1378                                  | AJ131501       | Homo Sapiens DNA sequence between two AML1 gene promoters, 6423 BP                                  | 5e-11   | 728831  | !!!! ALU SUBFAMILY J WARNING ENTRY   | 0.20    |
| 1379                                  | M27826         | Human endogenous retroviral protease mRNA, complete cds.  | 5e-11   | 88558   | retroviral proteinase-like protein - human   | 0.002   |
| 1380                                  | U23804         | Drosophila melanogaster putative GTP-binding regulatory protein beta chain (GPB) mRNA, partial cds. | 5e-11   | 2494916   | HYPOTHETICAL 55.2 KD TRP-ASP REPEATS CONTAINING PROTEIN T10F2.4 IN CHROMOSOME III protein; similar to G-Beta repeat region (Trp-Asp domains) of guanine nucleotide binding protein | 1e-30   |
| 1381                                  | Z22784         | M.musculus troponin I gene.   | 3e-11   | 3892202   | (AF072889) transcription repressor brain factor 2  | 0.053   |
| 1382                                  | AB007880       | Homo sapiens KIAA0420 mRNA, complete cds  | 2e-11   | <NONE>  | <NONE>   | <NONE>  |
| 1383                                  | AF020361       | 9 Homo sapiens BAX gene, exon 6, partial sequence   | 2e-11   | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1384   | L35600                                | Homo sapiens DNA sequence.   | 2e-11   | 1174952  | GLYCOPROTEIN D PRECURSOR gD [Bovine herpesvirus 1]             | 0.25    |
| 1385   | U21943                                | Human organic anion transporting polypeptide   | 2e-11   | 2738223  | (U95011) brain-specific organic anion transporter              | 9e-19   |
| 1386   | U90878                                | Homo sapiens carboxyl terminal LIM domain protein                                      | 2e-11   | 2996196  | (AF053367) carboxyl terminal LIM domain protein [Mus musculus] | 4e-23   |
| 1387   | U31929                                | Human orphan nuclear receptor (DAX1) gene, complete cds                                | 6e-12   | <NONE>   | <NONE>   | <NONE>  |
| 1388   | M25828                                | Human von Willebrand factor gene, exon 1, 2, and 3, and three Alu repetitive elements. | 6e-12   | <NONE>   | <NONE>   | <NONE>  |
| 1389   | AB020648                              | Homo sapiens mRNA for KIAA0841 protein, partial cds                                    | 3e-12   | <NONE>   | <NONE>   | <NONE>  |
| 1390   | Z15026                                | H.sapiens genes for tumor necrosis factor (Tnfa) and lymphotoxine (Tnfb)               | 2e-12   | <NONE>   | <NONE>   | <NONE>  |
| 1391   | L28101                                | Homo sapiens kallistatin (PI4) gene, exons 1-4, complete cds                           | 2e-12   | <NONE>   | <NONE>   | <NONE>  |
| 1392   | Z47046                                | Human cosmid QLL2C9 from Xq28  | 2e-12   | <NONE>   | <NONE>   | <NONE>  |
| 1393   | Z79007                                | H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA20E2                         | 2e-12   | 106322   | hypothetical protein (L1H 3' region) - human                   | 1.5     |
| 1394   | U34377                                | Human tyrosine kinase TXK (txk) gene, exon 13.   | 1e-12   | 151484   | (M55524) ORF 4; putative [Pseudomonas aeruginosa]              | 4.3     |



| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1395                                  | D70845         | Mus musculus apg-1 gene for novel member of heat shock protein 110, promoter region                   | 1e-12   | 113658  | ALKALINE PROTEINASE RECURSOR (ALP) precursor - fungus (Acremonium chrysogenum)                    | 3.5     |
| 1396                                  | M63978         | Human vascular endothelial growth factor gene, exon 8.  | 1e-12   | 3982737   | (AF069731) calmodulin-dependent protein kinase II beta M isoform [Rattus norvegicus]              | 0.083   |
| 1397                                  | U60266         | Homo sapiens lysosomal alpha-mannosidase (manB) mRNA, complete cds                                    | 8e-13   | <NONE>  | <NONE>  | <NONE>  |
| 1398                                  | Z68297         | Caenorhabditis elegans cosmid F11A10, complete sequence [Caenorhabditis elegans]                      | 7e-13   | 2393734   | (AC002542) similar to C. elegans F11A10.5; 80% similarity to Z68297 (PID:g1130619) [Homo sapiens] | 5e-34   |
| 1399                                  | Z68297         | Caenorhabditis elegans cosmid F11A10, complete sequence [Caenorhabditis elegans]                      | 7e-13   | 2393734   | (AC002542) similar to C. elegans F11A10.5; 80% similarity to Z68297 (PID:g1130619) [Homo sapiens] | 3e-38   |
| 1400                                  | Z68885         | Human DNA sequence from cosmid L21F12B, Huntington's Disease Region, chromosome 4p16.3, contains EST. | 6e-13   | <NONE>  | <NONE>  | <NONE>  |
| 1401                                  | X76104         | H.sapiens DAP-kinase mRNA   | 6e-13   | 2911154   | (AB007143) ZIP-kinase [Mus musculus]  | 0.007   |
| 1402                                  | Z78668         | H.sapiens flow-sorted chromosome 6 TaqI fragment. SC6pA13G4   | 5e-13   | 106322  | hypothetical protein (L2H 3' region) - human  | 2e-06   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1403                                  | L35600         | Homo sapiens DNA sequence.  | 3e-13   | 3184290   | (AC004136) hypothetical protein [Arabidopsis thaliana]   | 1.7     |
| 1404                                  | AF090452       | Cloning vector pKODT complete sequence  | 2e-13   | 3876730   | (Z49966) F35C11.4 [Caenorhabditis elegans]   | 7.8     |
| 1405                                  | D28126         | Human gene for ATP synthase alpha subunit, complete cds (exon 1 to 12)              | 2e-13   | 419481  | gag polyprotein - human endogenous virus S71   | 3.4     |
| 1406                                  | AF005219       | Homo sapiens transcription factor HOXD13  | 2e-13   | 2822166   | (AC004080) transcription factor HOXA13 [Homo sapiens]  | 5e-09   |
| 1407                                  | AB018301       | Homo sapiens mRNA for KIAA0758 protein, partial cds                                 | 2e-13   | 3882237   | (AB018301) KIAA0758 protein [Homo sapiens]   | 1e-23   |
| 1408                                  | D70845         | Mus musculus apg-1 gene for novel member of heat shock protein 110, promoter region | 1e-13   | 113658  | ALKALINE PROTEINASE PRECURSOR (ALP) precursor - fungus (Acremonium chrysogenum)  | 3.1     |
| 1409                                  | AG000691       | Homo sapiens genomic DNA, 21q region, clone: T171BG33                               | 8e-14   | 930045  | (X15332) alpha-1 (III) collagen [Homo sapiens]   | 3e-04   |
| 1410                                  | D30785         | Mouse mRNA for neuropsin, complete cds  | 8e-14   | 3559978   | (AJ005641) serine protease [Rattus rattus]   | 2e-12   |
| 1411                                  | U32710         | Haemophilus influenzae Rd section 25 of 163 of the complete genome                  | 8e-14   | 4106673   | (AL035064) queuine trna-ribosyltransferase [Schizosaccharomyces pombe]   | 2e-38   |
| 1412                                  | AG000886       | Homo sapiens genomic DNA, 21q region, clone: 64E11X19                               | 7e-14   | 1363925   | hypothetical protein 2 - North American opossum (fragment) >gi 897721 (Z48955) ORF-2, putative RT [Didelphis virginiana] | 1.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1413   | Z62664                                | H.sapiens CpG DNA, clone 71d11, forward read cpg71d11.ft1a.                 | 7e-14   | 3953461  | (AC002328) F20N2.6 [Arabidopsis thaliana]  | 0.0850  |
| 1414   | AB014532                              | Homo sapiens mRNA for KIAA0632 protein, partial cds                         | 7e-14   | 113668   | !!!! ALU CLASS C WARNING ENTRY !!!!  | 0.040   |
| 1415   | Z96478                                | H.sapiens telomeric DNA sequence, clone 20PTEL004, read 20PTELOO004.seq     | 7e-14   | 2981631  | (AB012223) ORF2 [Canis familiaris]   | 2e-04   |
| 1416   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                    | 4e-14   | <NONE>   | <NONE>   | <NONE>  |
| 1417   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds   | 4e-14   | <NONE>   | <NONE>   | <NONE>  |
| 1418   | AF033349                              | Homo sapiens MLL gene breakpoint cluster region, intron 1, partial sequence | 3e-14   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY   | 9.3     |
| 1419   | AC001526                              | Homo sapiens (subclone 4_f6 from P1 H54) DNA sequence                       | 3e-14   | 99861  | extensin - almond >gi 20420 (X65718) extensin  | 9.2     |
| 1420   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds   | 3e-14   | 728832   | !!!! ALU SUBFAMILY SB WARNING ENTRY  | 0.15    |
| 1421   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                    | 2e-14   | 3913573  | EPHRIN-A2 PRECURSOR (EPH-RELATED RECEPTOR TYROSINE KINASE LIGAND 6) (LERK-6) sapiens] >gi 2924761 (AC004258) EPL6_HUMAN [Homo sapiens] | 8.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1422   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 9e-15   | 119040   | ETB PROTEIN, SMALL T-ANTIGEN (E1B 19K)<br>>gi 74142 pir  Q1AD25 early E1B 21K protein II - human adenovirus 5>gi 58489 (X02996) mRNA 5 first reading frame [Human adenovirus type 5] adenovirus type 5] >gi 209797 (J01969_21kD protein | 1.5     |
| 1423   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 8e-15   | 477102   | transcription factor GATA-4, retinoic acid-inducible - mouse >gi 293345 (M98339) GATA-binding transcription factor [Mus musculus]   | 0.57    |
| 1424   | AB012223                              | Canis familiaris LINE 1 element ORF2 mRNA, complete cds                   | 8e-15   | 92385  | hypothetical protein - rate (fragment)  | 0.003   |
| 1425   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                  | 3e-15   | <NONE>   | <NONE>  | <NONE>  |
| 1426   | X12433                                | Human pHS1-2 mRNA with ORF homologous to membrane receptor proteins       | 3e-15   | 422532   | collagen alpha 3(IV) chain - sea urchin   | 8.9     |
| 1427   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 3e-15   | 1353143  | PROBABLE NUCLEAR HORMONE RECEPTOR E02H1.7 >gi 3875431 gnl PID e 1344980 (Z47075) similar to Zinc finger, C4 type (two domains) [Caenorhabditis elegans]   | 5.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 1428   | X69651                                | Human DNA sequence from cosmid L75B9, Huntington's Disease Region, chromosome 4p16.3           | 3e-15   | 403460   | (L24521) transformation-related protein [Homo sapiens]   | 0.60    |
| 1429   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                      | 2e-15   | 108750   | Ig heavy chain precursor (B/MT.4A.17.H5.A5) - bovine >gi 440 (X62916) anti-testosterone antibody [Bos taurus]  | 1.1     |
| 1430   | X83299                                | H.sapiens SMA3 mRNA  | 2e-15   | 671530   | (X83299) SMA3 gene product [Homo sapiens]  | 0.32    |
| 1431   | U01877                                | Human p300 protein mRNA, complete cds. > :: gb I62297 I62297 Sequence 1 from patent US 5658784 | 2e-15   | 3024341  | E1A-ASSOCIATED PROTEIN P300  | 0.019   |
| 1432   | X16516                                | Mouse MHC (Qa) Q2 k gene for class I antigen, exons 4-8  | 1e-15   | 2496897  | HYPOTHETICAL 45.1 KD PROTEIN C16C10.6 IN CHROMOSOME III >gi 3874384 gnl PID e 1344078 EST EMBL:C08256 comes from this gene; cDNA EST EMBL:C09941 comes from this gene; cDNA EST yk340a10.3 comes from this gene; cDNA EST yk340a10.5 comes from this gene [Ca... | 7e-08   |
| 1433   | M74165                                | Chicken tensin mRNA, complete cds.   | 1e-15   | 283920   | tensin - chicken >gi 212752 (M74165) tensin  | 2e-19   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1434                                  | X71893         | H.sapiens gene for immunoglobulin kappa light chain variable region O4 and O5     | 9e-16   | <NONE>  | <NONE>   | <NONE>  |
| 1435                                  | U05227         | Human Rar protein mRNA, complete cds.   | 9e-16   | 3036779   | (Z84479) match: multiple proteins; match: O00407 Q12829 P22127 P36861 Q40219; match: P70550 Q41022 P22125 Q08155 P35286; match: P51148 P51147 P35293 P36861 P35289; match: P35284 Q40217 P51152 P51157 P51158; match: Q41022 | 3e-06   |
| 1436                                  | M23404         | Chicken erythrocyte anion transport protein (band3) mRNA, complete cds.           | 9e-16   | 726403  | (U23175) similar to anion exchange protein [Caenorhabditis elegans]  | 1e-28   |
| 1437                                  | X16145         | Rat mRNA for liver $\alpha$ -L-Fucosidase (EC 3.2.1.51)                           | 9e-16   | 67502   | $\alpha$ -L-fucosidase (EC 3.2.1.51) 1 precursor, tissue - human >gi 178409 (M29877) $\alpha$ -L-fucosidase precursor (EC 3.2.1.5) [Homo sapiens]  | 2e-29   |
| 1438                                  | AF012899       | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds         | 8e-16   | <NONE>  | <NONE>   | <NONE>  |
| 1439                                  | AF076981       | Mus musculus brain mitochondrial carrier protein BMCP1 (Bmcp1) mRNA, complete cds | 8e-16   | 3851540   | (AF078544) brain mitochondrial carrier protein-1 [Homo sapiens]  | 2e-13   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1440   | Z54349                                | H.sapiens MN/CA9 GENE  | 5e-16   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY                                       | 0.002   |
| 1441   | AF077003                              | Mus musculus SH3 domain-containing adapter protein mRNA, complete cds                                  | 3e-16   | 309123   | (M35526) complement component C5D [Mus musculus]                         | 3.1     |
| 1442   | X64587                                | M.musculus mRNA for splicing factor U2AF (65 kD)   | 3e-16   | 2143767  | glycoprotein - rate >gi 986943 (L08134) glycoprotein [Rattus norvegicus] | 0.003   |
| 1443   | AB014561                              | Homo sapiens mRNA for KIAA0661 protein, complete cds   | 3e-16   | 3327136  | (AB014561) KIAA0661 protein [Homo sapiens]                               | 1e-20   |
| 1444   | Z73987                                | Human DNA sequence from cosmid N120B6 on chromosome 22 Contains ESTs, complete sequence [Homo sapiens] | 1e-16   | <NONE>   | <NONE>   | <NONE>  |
| 1445   | M58318                                | Homo sapiens ala gene.   | 1e-16   | <NONE>   | <NONE>   | <NONE>  |
| 1446   | U44103                                | Human small GTP binding protein Rab9 mRNA, complete cds  | 1e-16   | 1552584  | (Z80233) hypothetical protein Rv0029                                     | 1.3     |
| 1447   | AB014561                              | Homo sapiens mRNA for KIAA0661 protein, complete cds   | 9e-17   | 3327136  | (AB014561) KIAA0661 protein [Homo sapiens]                               | 2e-20   |
| 1448   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 1e-17   | <NONE>   | <NONE>   | <NONE>  |
| 1449   | M76762                                | Mus musculus ribosomal protein (Ke 3) gene, exons 1 to 5, and complete cds.                            | 1e-17   | 1073048  | pupR protein - Pseudomonas putida >gi 525260                             | 0.36    |
| 1450   | D50561                                | Human DNA, replication enhancing element (REE1)  | 4e-18   | 126295   | LINE-1 REVERSE TRANSCRIPTASE HOMOLOG                                     | 0.78    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION   | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 1451   | D16431                                | Human mRNA for hepatoma-derived growth factor, complete cds   | 4e-18   | 3242079  | (AJ006984) proline-rich protein  | 0.018   |
| 1452   | AF088983                              | Mus musculus heat shock protein hsp40-3 mRNA, complete cds    | 4e-18   | 3873707  | (Z73102) Similarity to B.subtilis DNAJ protein (SW:DNAJ_BACSU); cDNA EST yk437a1.5 comes from this gene [Caenorhabditis elegans] | 9e-25   |
| 1453   | U60205                                | Human methyl sterol oxidase (ERG25) mRNA, complete cds        | 3e-18   | <NONE>   | <NONE>   | <NONE>  |
| 1454   | AF038177                              | Homo sapiens clone 23899 mRNA sequence                        | 1e-18   | 1360775  | G protein-coupled receptor 74 - equine herpesvirus 2 >gi 695246 (U20824) G protein-coupled receptor [Equine herpesvirus 2]       | 5.1     |
| 1455   | AB014561                              | Homo sapiens mRNA for KIAA0661 protein, complete cds          | 1e-18   | 3327136  | (AB014561) KIAA0661 protein [Homo sapiens]   | 1e-21   |
| 1456   | AB014561                              | Homo sapiens mRNA for KIAA0661 protein, complete cds          | 1e-18   | 3327136  | (AB014561) KIAA0661 protein [Homo sapiens]   | 1e-22   |
| 1457   | U34374                                | Human tyrosine kinase TXK (txk) gene, exons 9 and 10.         | 1e-19   | <NONE>   | <NONE>   | <NONE>  |
| 1458   | AB006969                              | Homo sapiens hGAA1 mRNA, complete cds                         | 1e-19   | 4151809  | (Af102855) synaptic SAPAP-interacting protein Synamon  | 0.19    |
| 1459   | AB002293                              | Human mRNA for KIAA0295 gene, partial cds                     | 1e-19   | 2224531  | (AB002293) KIAA0295 [Homo sapiens]   | 6e-17   |
| 1460   | Z59664                                | H.sapiens CpG DNA, clone 168t9, reverse read cpg 168f9.rt la. | 5e-20   | 3880251  | (Z82055) predicted using Genefinder  | 6.5     |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1461                                  | M73837         | Human modulator recognition factor 2 (MRF-2) mRNA, complete cds.   | 5e-20   | 284313  | modulator recognition factor 2 - human factor 2 [Homo sapiens]   | 0.019   |
| 1462                                  | U24267         | Human pyrroline-5-carboxylate dehydrogenase                        | 5e-20   | 2506350   | DELTA-T-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE PRECURSOR (P5C DEHYDROGENASE _>gi 1353248 sapiens] >gi 1353250 (U24267) pyrroline 5-carboxylate dehydrogenase [Homo sapiens] >gi 1589585 prf 2211355A Delta 1-pyrroline-5-carboxylate dehydrogenase [Homo sapiens] | 5e-04   |
| 1463                                  | U13262         | Mus musculus myelin gene expression factor                         | 4e-20   | 536926  | (U13262) myelin gene expression factor [Mus musculus]  | 3e-07   |
| 1464                                  | U13262         | Mus musculus myelin gene expression factor                         | 4e-20   | 3126878   | (AF061832) M4 protein deletion mutant [Homo sapiens]   | 1e-08   |
| 1465                                  | Z61239         | H.sapiens CpG DNA, clone 48f10, forward read cpg48f10.ft1a.        | 4e-20   | 1669601   | (D88747) AR401 [Arabidopsis thaliana]  | 8e-19   |
| 1466                                  | U89915         | Mus musculus junctional adhesion molecule (Jam) mRNA, complete cds | 1e-20   | 3462455   | (U89915) junctional adhesion molecule [Mus musculus]   | 7e-11   |
| 1467                                  | AF029071       | Gallus gallus p52 proapoptotic protein mRNA, complete cds          | 7e-22   | 2599492   | (AF029071) p52 proapoptotic protein [Gallus gallus]  | 1e-15   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1468                                  | M25636         | Figure 4 Nucleotide sequence of the pKS36 1.797 kb insert.      | 6e-22   | 1196398   | (M21305) unknown protein [Homo sapiens]   | 0.65    |
| 1469                                  | AB020655       | Homo sapiens mRNA for KIAA0848 protein, complete cds            | 6e-22   | 4240325   | (AB020725) KIAA0918 protein [Homo sapiens]  | 1e-19   |
| 1470                                  | S80935         | chorionic gonadotropin beta 1 (CG beta 1) subunit               | 5e-22   | 115310  | PROCOLLAGEN ALPHA I (IV) CHAIN PRECURSOR<br>>gi 84917 pir  A31893 collagen alpha I (IV) chain precursor - fruit fly (Drosophila melanogaster) melanogaster]<br>>gi 157078 (M96575) type IV collagen procollagen [Drosophila melanogaster] | 0.027   |
| 1471                                  | AF053066       | Homo sapiens microsatellite D5S2926 sequence                    | 2e-22   | 728831  | !!!! ALU SUBFAMILY J WARNING ENTRY  | 3e-04   |
| 1472                                  | U55177         | Danio rerio carbonic anhydrase homolog CAH-Z mRNA, complete cds | 2e-22   | 3123190   | CARBONIC ANHYDRASE (CARBONATE DEHYDRATASE)<br>>gi 2576335 (U55177) CAH-Z [Danio rerio]  | 5e-14   |
| 1473                                  | AF064250       | Gallus gallus ubiquitin specific protease 66                    | 2e-22   | 2736064   | (AF016107) ubiquitin specific protease 41 [Gallus gallus]   | 7e-37   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1474                                  | AF030880       | Homo sapiens pendrin (PDS) mRNA, complete cds                                      | 2e-22   | 729367  | DRA PROTEIN (DOWN-REGULATED IN ADENOMA)<br>>gi 2135020 pir  A47456 down-regulated in adenoma (DRA) - human >gi 291964 (L02785) Nuclear localization signal at AA 569-573, 576-580, 579-583; acidic transcr. activ. domain 620-640; homeobox motif 653-676 [Homo sapiens] | 4e-53   |
| 1475                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds   | 6e-23   | <NONE>  | <NONE>   | <NONE>  |
| 1476                                  | X57398         | Human mRNA for pM5 protein   | 3e-23   | 107350  | Pm5 protein - human<br>>gi 1335273 gnl PID e36241  | 1e-04   |
| 1477                                  | AB010998       | Rattus norvegicus PAD-R11 mRNA for Peptidylarginine deiminase type I, complete cds | 2e-23   | <NONE>  | <NONE>   | <NONE>  |
| 1478                                  | D10871         | Human h NAT allele 2-2 gene for arylamine N-acetyltransferase                      | 2e-23   | 171200  | (J04734) CDC6 protein [Saccharomyces cerevisiae]   | 9.8     |
| 1479                                  | D10871         | Human h NAT allele 2-2 gene for arylamine N-acetyltransferase                      | 2e-23   | 171200  | (J04734) CDC6 protein [Saccharomyces cerevisiae]   | 8.3     |
| 1480                                  | AF024541       | Homo sapiens MLL-AF4 fusion protein mRNA, partial cds                              | 2e-23   | 2136142   | serine/proline-rich FEL protein, splice form 1 - human   | 1e-20   |
| 1481                                  | L13773         | Human AF-4 mRNA, complete cds.   | 2e-23   | 3063962   | (AF031404) MLL-AF4 fusion protein [Homo sapiens]   | 1e-20   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1482   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 8e-24   | <NONE>   | <NONE>   | <NONE>  |
| 1483   | U75467                                | Drosophila melanogaster Rga and Atu genes, complete cds                             | 8e-24   | 1658503  | (U75467) Atu [Drosophila melanogaster]   | 2e-37   |
| 1484   | D17076                                | Human HepG2 partial cDNA, clone hmd5a09m5   | 7e-24   | <NONE>   | <NONE>   | <NONE>  |
| 1485   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 7e-24   | 1169643  | FMRFAMIDE-RELATED NEUROPEPTIDES PRECURSOR >gi 416208 (U03137) neuropeptide precursor FMRFamide-related peptide [Lymnaea stagnalis] | 7e-10   |
| 1486   | M11167                                | Human 28S ribosomal RNA gene.   | 2e-24   | 3875481  | (Z81054) predicted using Genefinder; Similarity to UDP-glucoronosyltransferases  | 5.1     |
| 1487   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 2e-24   | 549173   | USP1 PROTEIN PRECURSOR >gi 169623  | 1.2     |
| 1488   | AB003468                              | Cloning vector pAP2neo DNA, complete sequence                                       | 2e-24   | 987050   | (X65335) lacZ gene product [unidentified cloning vector]   | 0.058   |
| 1489   | X03541                                | Human mRNA of trk oncogene > :: gb I96186 I96186 Sequence 23 from patent US 5734039 | 2e-24   | 325465   | (M74509) [Human endogenous retrovirus type C oncovirus sequence.], gene product [Homo sapiens]                                     | 3e-04   |
| 1490   | L81652                                | Homo sapiens (subclone 2_g11 from P1 H43) DNA sequence                              | 2e-24   | 225047   | reverse transcriptase related protein [Homo sapiens]   | 4e-12   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1491                                  | U95760         | Drosophila melanogaster strawberry notch (sno) mRNA, complete cds          | 2e-24   | 2078282   | (U95760) Sno [Drosophila melanogaster]  | 2e-41   |
| 1492                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds                                   | 8e-25   | 2623773   | (AF004835) tyrocidine synthetase 3 [Brevibacillus brevis]   | 8.6     |
| 1493                                  | AB002405       | Homo sapiens mRNA for LAK-4p, complete cds                                 | 8e-25   | 2496822   | HYPOTHETICAL 127.3 KD PROTEIN B0416.1 IN CHROMOSOME X >gi 746502 (U23516) B0416.1 gene product [Caenorhabditis elegans] | 9e-11   |
| 1494                                  | K03002         | Human mRNA from chromosome 15 gene with homology to MHC-HLA-SB-1 intron A. | 8e-25   | 1514614   | (X92842) nuclear protein [Mus musculus]   | 1e-13   |
| 1495                                  | U61232         | Human tubulin-folding cofactor E mRNA, complete cds                        | 7e-25   | 1465772   | (U61232) cofactor E [Homo sapiens]  | 2e-05   |
| 1496                                  | U10245         | Arabidopsis thaliana Col-O putative RNA helicase A mRNA, complete cds.     | 5e-25   | 1353239   | (U10245) putative RNA helicase A [Arabidopsis thaliana]   | 1e-37   |
| 1497                                  | X89211         | H.sapiens DNA for endogenous retroviral like element                       | 3e-25   | 2065210   | (Y12713) Pro-Pol-dUTPase polypeptide  | 5e-06   |
| 1498                                  | L81652         | Homo sapiens (subclone 2_g11 from P1 H43) DNA sequence                     | 3e-25   | 2072961   | (U93568) putative p150 [Homo sapiens]   | 5e-16   |
| 1499                                  | X82895         | H.sapiens mRNA for DLG2  | 2e-25   | 2497511   | MAGUK P55 SUBFAMILY MEMBER 2 (MPP2 PROTEIN) (DISCS. LARGE HOMOLOG 2)  | 1e-34   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION   | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 1500   | M36654                                | Mouse homeo box 2.6 (Hox-2.6) mRNA, complete cds.       | 9e-26   | 3323169  | (AE001255) T. pallidum predicted coding region TP0854   | 1.9     |
| 1501   | L36315                                | Mus musculus (Clone pMLZ-1) zinc finger protein         | 9e-26   | 1806134  | (Z67747) zinc finger protein [Mus musculus]   | 4e-05   |
| 1502   | AB018281                              | Homo sapiens mRNA for KIAA0738 protein, complete cds    | 9e-26   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY  | 1e-07   |
| 1503   | AF017433                              | Homo sapiens putative transcription factor CR53         | 9e-26   | 3219985  | ZINC FINGER PROTEIN ZFP-29  | 1e-17   |
| 1504   | AC001225                              | Homo sapiens (subclone 2_e6 from BAC H94) DNA sequence  | 8e-26   | 2653713  | (U91823) small S protein [Hepatitis B virus]  | 4.3     |
| 1505   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                | 8e-26   | 283446   | cyteine-rich surface antigen 72, CRP72 - Giardia lamblia (fragment)   | 3.4     |
| 1506   | X94912                                | H.sapiens Pr22 gene                                     | 3e-26   | 728837   | !!!! ALU SUBFAMILY SQ WARNING ENTRY   | 4e-09   |
| 1507   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                | 2e-26   | <NONE>   | <NONE>  | <NONE>  |
| 1508   | U44103                                | Human small GTP binding protein Rab9 mRNA, complete cds | 1e-26   | 3327038  | (AB014512) KIAA0612 protein [Homo sapiens]  | 8.7     |
| 1509   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                | 9e-27   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 0.14    |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1510                                  | AG001212       | Homo sapiens genomic DNA, 21q region, clone: 9H11N46                 | 9e-27   | 126296  | LINE-1 REVERSE TRANSCRIPTASE HOMOLOG protein [Nycticebus coucang]                 | 0.012   |
| 1511                                  | AF027131       | Mus musculus mucin glycoprotein MUC3 mRNA, partial cds               | 9e-27   | 2589172   | (U76551) mucin Mus3 [Rattus norvegicus]   | 2e-14   |
| 1512                                  | U49057         | Rattus norvegicus CTD-binding SR-like protein rA9 mRNA, complete cds | 5e-27   | 1438534   | (U49057) rA9 [Rattus norvegicus]  | 1e-04   |
| 1513                                  | J03764         | Human, plasminogen activator inhibitor-1 gene, exons 2 to 9.         | 3e-27   | <NONE>  | <NONE>  | <NONE>  |
| 1514                                  | Z78160         | M.musculus partial cochlear mRNA (clone 28D2)                        | 3e-27   | 1490362   | (Z78160) unknown [Mus musculus]   | 2e-05   |
| 1515                                  | Z64210         | H.sapiens CpG DNA, clone 99b4, reverse read cpg99b4.rta.             | 3e-27   | 2257538   | (AB004538) LIPOIC ACID SYNTHETASE PRECURSOR (LIP-SYN) [Schizosaccharomyces pombe] | 1e-06   |
| 1516                                  | L35659         | Homo sapiens (subclone H8 6_h6 from P1 35 H5 C8) DNA sequence.       | 1e-27   | <NONE>  | <NONE>  | <NONE>  |
| 1517                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds                             | 1e-27   | 1644471   | (U72686) odorant receptor 4 [Danio rerio]   | 7.5     |
| 1518                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds                             | 1e-27   | 2738388   | (AF003534) hypothetical protein 004L [Chilo iridescent virus]                     | 6.7     |
| 1519                                  | AB009271       | Homo sapiens gene for BCNT, partial cds                              | 1e-27   | 3880909   | (AL032636) Y40B1B.3 [Caenorhabditis elegans]                                      | 4.6     |
| 1520                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds                             | 1e-27   | 2133579   | spermatophorin Sp23 - yellow mealworm molitor]                                    | 0.85    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1521   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                     | 1e-27   | 121805   | ENDOGLUCANASE A PRECURSOR   | 0.58    |
| 1522   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                     | 1e-27   | 3722000  | (AF035323) survival motor neuron protein [Bos taurus]   | 0.10    |
| 1523   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                     | 1e-27   | 3328188  | (AF074902) laminin alpha chain [Caenorhabditis elegans]   | 0.083   |
| 1524   | AF074382                              | Homo sapiens Ikb kinase gamma subunit                        | 1e-27   | 3641280  | (AF074382) Ikb kinase gamma subunit [Homo sapiens]  | 0.041   |
| 1525   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                     | 1e-27   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExtI) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 6e-04   |
| 1526   | L78778                                | Homo sapiens (subclone 2_e10 from P1 H49) DNA sequence       | 1e-27   | 225047   | reverse transcriptase related protein [Homo sapiens]  | 2e-09   |
| 1527   | L03427                                | Human zinc finger protein basonuclin mRNA, complete cds.     | 1e-27   | 1488275  | (U59694) zinc finger protein basonuclin [Homo sapiens]  | 9e-22   |
| 1528   | U09954                                | Human ribosomal protein L9 gene. 5; region and complete cds. | 4e-28   | 2257538  | (AB004538) LIPOIC ACID SYNTHETASE PRECURSOR (LIP-SYN) [Schizosaccharomyces pombe]   | 2e-04   |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1529                                  | Z64210         | H.sapiens CpG DNA,<br>clone 99b4, reverse<br>read cpg99b4.r1la.                          | 4e-28   | 3878570   | (Z46381) similar to<br>lipoic acid synthase;<br>cDNA EST yk283b6.3<br>comes from this gene;<br>cDNA EST yk283b6.5<br>comes from this gene;<br>cDNA EST yk472f5.3<br>comes from this gene;<br>cDNA EST yk472f5.5<br>comes from this gene;<br>cDNA EST<br>yk476e7.3... | 7e-11   |
| 1530                                  | U55177         | Danio rerio carbonic<br>anhydrase homolog<br>CAH-Z mRNA,<br>complete cds                 | 4e-28   | 3123190   | CARBONIC<br>ANHYDRASE<br>(CARBONATE<br>DEHYDRATASE)<br>>gi 2576335 (U55177)<br>CAH-Z [Danio rerio]   | 5e-21   |
| 1531                                  | D43682         | Human mRNA for<br>very-long-chain acyl-<br>CoA dehydrogenase<br>(VLCAD), complete<br>cds | 4e-28   | 1351839   | ACYL-COA<br>DEHYDROGENASE,<br>VERY-LONG-CHAIN<br>SPECIFIC<br>PRECURSOR<br>(VLCAD) >gi 930358<br>taurus]  | 3e-27   |
| 1532                                  | AF016591       | Homo sapiens survival<br>motor neuron<br>pseudogene, complete<br>sequence                | 3e-28   | 728831  | !!!! ALU<br>SUBFAMILY SB<br>WARNING ENTRY  | 3e-08   |
| 1533                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds   | 2e-28   | 728832  | !!!! ALU<br>SUBFAMILY SB<br>WARNING ENTRY  | 2.5     |
| 1534                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds   | 2e-28   | 118588  | DEHYDRIN DHN3<br>>gi 100035 pir  S1813<br>9 dehydrin DHN3 -<br>garden pea >gi 20709<br>(X63063) pea<br>dehydrin DHN3<br>[Pisum sativum]  | 0.004   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1535                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 1169643   | FMRFAMIDE-<br>RELATED<br>NEUROPEPTIDES<br>PRECURSOR<br>>gi 416208 (U03137)<br>neuropeptide<br>precursor<br>FMRFamide-related<br>peptide [Lymnaea<br>stagnalis]   | 6e-04   |
| 1536                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 9e-05   |
| 1537                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 2e-06   |
| 1538                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 2e-06   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1539                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 1e-09   |
| 1540                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-10   |
| 1541                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 1e-11   |
| 1542                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 2e-28   | 3157926   | (AC002131) Strong<br>similarity to extensin-<br>like protein gb Z34465<br>from Zea mays.<br>[Arabidopsis thaliana]   | 8e-12   |
| 1543                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>   | <NONE>  |
| 1544                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|---------------------------------------|----------------|--|---------|---|-------------|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION | P VALUE |
| 1545                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1546                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1547                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1548                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1549                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1550                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1551                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1552                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1553                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1554                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1555                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1556                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |
| 1557                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | <NONE>  | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCES-SION                            | DESCRIPTION                              | P VALUE | ACCES-SION   | DESCRIPTION | P VALUE |
| 1558   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1559   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1560   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1561   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1562   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1563   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1564   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1565   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1566   | M87708                                | Human simple repeat polymorphism.        | 1e-28   | <NONE>   | <NONE>      | <NONE>  |
| 1567   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>   | <NONE>      | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1568                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 3924779   | (AL008585) similar to<br>rammin B; cDNA EST<br>yk450d8.5 comes<br>from this gene; cDNA<br>EST yk249a6.5 comes<br>from this gene; cDNA<br>ESTyk219a2.5 comes<br>from this gene; cDNA<br>EST yk355e4.5 comes<br>from this gene; cDNA<br>EST yk244f4.5 comes<br>fr...<br><br>>gi 3924881 gnl PID e<br>1354569 from this<br>gene; cDNA EST<br>yk219a2.5 comes from<br>this gene; cDNA EST<br>yk355e4.5 comes from<br>this gene; cDNA EST<br>yk244f4.5 comes<br>from... | 3.0     |
| 1569                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 1196943   | FMRFAMIDE-<br>RELATED<br>NEUROPEPTIDES<br>PRECURSOR<br>>gi 416208 (U03137)<br>neuropeptide<br>precursor<br>FMRFamide-related<br>peptide [Lymnaea<br>stagnalis]   | 0.66    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION                              | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 1570   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 3924779  | (AL008585) similar to rammin B, cDNA EST yk450d8.5 comes from this gene; cDNA EST yk249a6.5 comes from this gene; cDNA EST yk219a2.5 comes from this gene; cDNA EST yk355e4.5 comes from this gene; cDNA EST yk224f4.5 comes fr...<br><br>>gi 3924881 gnl PID e 1354569 from this gene; cDNA EST yk249a6.5 comes from this gene; cDNA EST yk219a2.5 comes from this gene; cDNA EST yk355e4.5 comes from this gene; cDNA EST yk224f4.5 comes from... | 0.65    |
| 1571   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 2133579  | spermatophorin Sp23 - yellow mealworm molitor]  | 0.49    |
| 1572   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 2133579  | spermatophorin Sp23 - yellow mealworm molitor]  | 4.49    |
| 1573   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 283446   | cyteine-rich surface antigen 72, CRP72 - Giardia lamblia (fragment)   | 0.45    |
| 1574   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 2498937  | SPERMATOPHORIN SP23 PRECURSOR mealworm >gi 161725 (M92928) structural protein   | 0.33    |
| 1575   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 1492050  | (U60315) MC107L [Molluscum contagiosum virus subtype 1]   | 0.18    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION                              | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1576   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 2133579  | spermatophorin Sp23 - yellow mealworm molitor]  | 0.088   |
| 1577   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 118588   | DEHYDRIN DHN3 >gi 100035 pir  S1813 9 dehydrin DHN3 - garden pea >gi 20709 (X63063) pea dehydrin DHN3 [Pisum sativum]   | 0.018   |
| 1578   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 118588   | DEHYDRIN DHN3 >gi 100035 pir  S1813 9 dehydrin DHN3 - garden pea >gi 20709 (X63063) pea dehydrin DHN3 [Pisum sativum]   | 0.016   |
| 1579   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 118588   | DEHYDRIN DHN3 >gi 100035 pir  S1813 9 dehydrin DHN3 - garden pea >gi 20709 (X63063) pea dehydrin DHN3 [Pisum sativum]   | 0.012   |
| 1580   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 0.010   |
| 1581   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 118588   | DEHYDRIN DHN3 >gi 100035 pir  S1813 9 dehydrin DHN3 - garden pea >gi 20709 (X63063) pea dehydrin DHN3 [Pisum sativum]   | 0.002   |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1582                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 1169643   | FMRFAMIDE-<br>RELATED<br>NEUROPEPTIDES<br>PRECURSOR<br>>gi 416208 (U03137)<br>neuropeptide<br>precursor<br>FMRFamide-related<br>peptide [Lymnaea<br>stagnalis]   | 0.002   |
| 1583                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 0.002   |
| 1584                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 118588  | DEHYDRIN DHN3<br>>gi 100035 pir  S1813<br>9 dehydrin DHN3 -<br>garden pea >gi 20709<br>(X63063) pea<br>dehydrin DHN3<br>[Pisum sativum]  | 0.002   |
| 1585                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 0.002   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1586                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 118588  | DEHYDRIN DHN3<br>>gi 100035 pir  S1813<br>9 dehydrin DHN3 -<br>garden pea >gi 20709<br>(X63063) pea<br>dehydrin DHN3<br>[Pisum sativum]  | 0.001   |
| 1587                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 0.001   |
| 1588                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 6e-04   |
| 1589                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-04   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1590                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-04   |
| 1591                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 118588  | DEHYDRIN DHN3<br>>gi 100035 pir  S1813<br>9 dehydrin DHN3 -<br>garden pea >gi 20709<br>(X63063) pea<br>dehydrin DHN3<br>[Pisum sativum]  | 2e-04   |
| 1592                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 2e-04   |
| 1593                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-05   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1594                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-05   |
| 1595                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 1e-05   |
| 1596                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 1e-05   |
| 1597                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990)Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana]  | 9e-06   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1598                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana]  | 6e-06   |
| 1599                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990)Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana]   | 5e-06   |
| 1600                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 544357  | RNA-BINDING<br>PROTEIN FUS/TLS<br>protein [human.<br>Peptide, 526 aa]<br>[Homo sapiens]   | 4e-06   |
| 1601                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extension<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 2e-06   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1602                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 2e-06   |
| 1603                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 9e-07   |
| 1604                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 8e-07   |
| 1605                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 1169643   | FMRFAMIDE-<br>RELATED<br>NEUROPEPTIDES<br>PRECURSOR<br>>gi 416208 (U03137)<br>neuropeptide<br>precursor<br>FMRFamide-related<br>peptide [Lymnaea<br>stagnalis]   | 7e-07   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1606                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 6e-07   |
| 1607                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-07   |
| 1608                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 3e-07   |
| 1609                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 1e-07   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1610                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 1e-07   |
| 1611                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 7e-08   |
| 1612                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 2e-08   |
| 1613                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 6e-09   |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1614                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-09   |
| 1615                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 4e-09   |
| 1616                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 7e-10   |
| 1617                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 6e-10   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION                                    | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1618                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-10   |
| 1619                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 4e-10   |
| 1620                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 2e-10   |
| 1621                                  | AF100694       | Mus musculus<br>Pontin52 mRNA,<br>complete cds | 1e-28   | 4056454   | (AC005990) Contains<br>repeated region with<br>similarity to<br>gb U43627 extensin<br>(atExt1) gene from<br>Arabidopsis thaliana.<br>ESTs gb Z34165 and<br>gb Z18788 come from<br>this gene.<br>[Arabidopsis thaliana] | 5e-11   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION                                      | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1622   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds         | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 2e-12   |
| 1623   | AF032896                              | Petromyzon marinus polyadenylate binding protein | 1e-28   | 1082703  | polyadenylate binding protein II human  | 2e-27   |
| 1624   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds         | 9e-29   | 118588   | DEHYDRIN DHN3 >gi 100035 pir  S1813 9 dehydrin DHN3 - garden pea >gi 20709 (X63063) pea dehydrin DHN3 [Pisum sativum]   | 0.013   |
| 1625   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds         | 9e-29   | 2133579  | spermatophorin Sp23 - yellow mealworm molitor]  | 6e-04   |
| 1626   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds         | 9e-29   | 3876465  | (Z81071) predicted using Genefinder; Similarity to Human small nuclear ribonucleoprotein E cDNA EST yk375g7.5 comes from this gene; cDNA EST yk435f5.3 comes from this gen...               | 9e-06   |
| 1627   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds         | 8e-29   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 2e-06   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1628                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds                                    | 4e-29   | 728883  | ADP-RIBOSYLATION Factor 3 fruit fly (Drosophila melanogaster) >gi 507234 (L25063) ADP ribosylation factor 3 [Drosophila melanogaster]   | 0.016   |
| 1629                                  | AF100694       | Mus musculus Pontin 52 mRNA, complete cds                                   | 4e-29   | 544357  | RNA-BINDING PROTEIN FUS/TLS protein [human, Peptide, 526aa] [Homo sapiens]  | 2e-07   |
| 1630                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds                                    | 4e-29   | 4056454   | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 1e-08   |
| 1631                                  | D43682         | Human mRNA for very-long-chain acyl-CoA dehydrogenase (VLCAD), complete cds | 4e-29   | 1168287   | ACYL-COA DEHYDROGENASE, VERY-LONG-CHAIN SPECIFIC PRECURSOR (VLCAD) dehydrogenase precursor - rat Acyl-CoA dehydrogenase [Rattus norvegicus]   | 6e-37   |
| 1632                                  | Y07660         | M. tuberculosis accBC gene  | 4e-29   | 2113935   | (Z95556) accD1 [Mycobacterium tuberculosis]   | 3e-47   |
| 1633                                  | X55367         | Human alpha-satellite DNA from clone pTRA-2.                                | 1e-29   | <NONE>  | <NONE>  | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1634                                  | L81866         | Homo sapiens (subclone 1_fl from P1 H54) DNA sequence                 | 1e-29   | <NONE>  | <NONE>  | <NONE>  |
| 1635                                  | S75940         | {Alu repeats, clone 52H10} [human, colonic mucosa, Genomic, 943 nt]   | 1e-29   | 728831  | !!!! ALU SUBFAMILY J WARNING ENTRY  | 1e-07   |
| 1636                                  | AB001907       | Homo sapiens PACE4 gene, exon 13                                      | 1e-29   | 728831  | !!!! ALU SUBFAMILY J WARNING ENTRY  | 2e-09   |
| 1637                                  | AF077003       | Mus musculus SH3 domain-containing adaptor protein mRNA, complete cds | 5e-30   | <NONE>  | <NONE>  | <NONE>  |
| 1638                                  | AF100694       | Mus musculus Pontin52 mRNA, complete cds                              | 4e-30   | 40564654  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 3e-10   |
| 1639                                  | M27072         | Xenopus laevis poly(A)-binding protein (ABP-EF) mRNA, complete cds.   | 4e-30   | 1352709   | POLYADENYLATE-BINDING PROTEIN polyadenylate-binding protein - African clawed frog laevis]   | 5e-21   |
| 1640                                  | X58386         | B.taurus mRNA from bovine vacuolar ATPase subunit A                   | 2e-30   | 2773154   | (AF039573) abscisic acid- and stress-inducible protein  | 4.3     |
| 1641                                  | Y07660         | M. tuberculosis accBC gene  | 1e-30   | 2113935   | (Z95556) accD1 [Mycobacterium tuberculosis]   | 4e-47   |
| 1642                                  | AJ236940       | Sus scrofa mRNA for hypothetical protein (5': clone 7C4)              | 4e-31   | 4102021   | (AF007561) delta 6-desaturase [Borago officinalis]  | 7.4     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1643                                  | AF039400       | Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA, complete cds  | 2e-31   | 3721912   | (AB017156) gob-5 [Mus musculus]   | 7e-08   |
| 1644                                  | L77036         | Homo sapiens (subclone 5_d9 from P1 H19) DNA sequence.   | 1e-31   | 461663  | BOMBYXIN B-2 HOMOLOG PRECURSOR silkmoth >gi 217385 gnl PID d1003528 (D13924) Samia bombyxin homolog B-2 [Samia cynthia]   | 1.1     |
| 1645                                  | X61971         | H. sapiens mRNA for macropain subunit delta  | 1e-31   | 296734  | (X61971) macropain subunit delta [Homo sapiens]   | 3e-06   |
| 1646                                  | L00016         | Human mitochondrial trnas and partial proteins 4 & 5; histidyl-, seryl-, leucyl-trna genes; urf4 and urf5 (partial). | 5e-32   | 4056454   | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene [Arabidopsis thaliana]  | 0.002   |
| 1647                                  | M17887         | Human acidic ribosomal phosphoprotein P2 mRNA, complete cds.   | 5e-32   | 4056454   | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 1e-05   |
| 1659                                  | U53446         | Human mitogen-responsive phosphoprotein DOC-2 mRNA, complete cds.  | 6e-34   | 3395443   | (AC004683) putative ammonium transporter, 3' partial  | 4.7     |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1660                                  | AF013988       | Homo sapiens serine protease mRNA, complete cds                                | 4e-34   | 2507226   | PROTEIN-TYROSINE PHOSPHATASE EPSILON PRECURSOR (R-PTP-EPSILON) >gi 1439605 (U62387) protein tyrosine phosphatase-e [Mus musculus]  | 3.2     |
| 1661                                  | U53446         | Human mitogen-responsive phosphoprotein DOC-2 mRNA, complete cds.              | 2e-34   | 104757  | LEP100 protein precursor - chicken >gi 212254 gallus]  | 1.6     |
| 1662                                  | AJ233632       | Homo sapiens endogenous retroviral sequence ERV-L pol gene, clone ERV-L Human6 | 2e-34   | 3860513   | (AJ233597) reverse transcriptase [Mus famulus]   | 4e-10   |
| 1663                                  | AF086310       | Homo sapiens full length insert cDNA clone ZD51F08                             | 8e-35   | 2947070   | (AC002521) putative Ser/Thr protein kinase [Arabidopsis thaliana]  | 2.3     |
| 1664                                  | X17206         | Human mRNA for LLRep3  | 3e-35   | 730652  | 40S RIBOSOMAL PROTEIN S2 (STRINGS OF PEARLS PROTEIN) >gi 085158 pir  S5032 5 ribosomal protein S2 - fruit fly (Drosophila melanogaster) melanogaster]> gi 515972 (U01335) ribosomal protein S2 | 2e-10   |
| 1665                                  | AB011137       | Homo sapiens mRNA for KIAA0565 protein, complete cds                           | 3e-35   | 3043654   | (AB011137) KIAA0565 protein [Homo sapiens]   | 2e-16   |
| 1666                                  | U62801         | Human protease M mRNA, complete cds  | 2e-35   | 3929231   | (AF091247) potassium channel [Rattus norvegicus]   | 1.0     |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1667                                  | AF020760       | Homo sapiens serine protease (Omi) mRNA, complete cds                                   | 1e-35   | 2738915   | (AF020760) serine protease [Homo sapiens]   | 9e-14   |
| 1668                                  | Z93943         | Human DNA sequence from cosmid U235H3 on chromosome X                                   | 8e-36   | 1196432   | (M22333) unknown protein [Homo sapiens]   | 3e-10   |
| 1669                                  | X06778         | Rabbit 18S rRNA   | 7e-36   | 118588  | DEHYDRIN DHN3 >gi 100035 pir  S1813 9 dehydrin DHN3 - garden pea >gi 20709 (X63063) pea dehydrin DHN3 [Pisum sativum] | 0.011   |
| 1670                                  | AB007962       | Homo sapiens mRNA, chromosome 1 specific transcript KIAA0493                            | 3e-36   | 3329243   | (AE001350) hypothetical protein [Chlamydia trachomatis]   | 3.1     |
| 1671                                  | Z81014         | Human DNA sequence from cosmid U65A4, between markers DXS366 and DXS87 on chromosome X* | 3e-36   | 141108  | HYPOTHETICAL PROTEIN ORF-1137 mouse   | 0.038   |
| 1672                                  | Z81014         | Human DNA sequence from cosmid U65A4, between markers DXS366 and DXS87 on chromosome X* | 3e-36   | 198651  | (M29325) ORF1 [Mus musculus]  | 0.006   |
| 1673                                  | U49082         | Human transporter protein (g17) mRNA, complete cds                                      | 3e-36   | 1840045   | (U49082) transporter protein [Homo sapiens]   | 2e-15   |
| 1674                                  | J03133         | Human transcription factor SP1 mRNA, 3' end.  | 3e-36   | 477133  | HF-1 regulatory element binding protein - rat   | 2e-31   |
| 1675                                  | AB007934       | Homo sapiens mRNA for KIAA0465 protein, partial cds                                     | 1e-36   | 3413892   | (AB007934) KIAA0465 protein [Homo sapiens]  | 4e-37   |



| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1676                                  | M34857         | Mouse Hox - 2.5 mRNA.   | 9e-37   | 106296  | Homeotic protein Hox B9 - human (fragment)   | 0.15    |
| 1677                                  | L35657         | Homo sapiens (subclone H8 5_a10 from P1 35 H5 C8) DNA sequence.   | 9e-37   | 2072960   | (U93568) p40 [Homo sapiens]  | 3e-05   |
| 1678                                  | X80240         | H. sapiens Endogenous retrovirus HERV-KC4 DNA   | 8e-37   | 4185944   | (Y17833) env protein [Human endogenous retrovirus K]   | 1e-15   |
| 1679                                  | Z93943         | Human DNA sequence from cosmid U235H3 on chromosome X   | 9e-38   | 106322  | Hypothetical protein (L1H 3' region) - human   | 4e-13   |
| 1680                                  | X97303         | H. sapiens mRNA for Ptg-12 protein  | 4e-38   | 466044  | HYPOTHETICAL ZINC FINGER PROTEIN ZK686.4 IN CHROMOSOME III<br>>gi 630780 pir  S44909 ZK686.4 protein - Caenorhabditis elegans >gi 304346 (L17337) coded for by C. elegans cDNAs GenBank:M88869 and T01933; putative [Caenorhabditis elegans] | 3e-37   |
| 1681                                  | Y08999         | H. sapiens mRNA for Sop2p-like protein  | 3e-38   | 3334339   | SOP2-LIKE PROTEIN  | 5e-06   |
| 1682                                  | Z62887         | H. sapiens CpG DNA, clone 74g6, forward read cpg74g6.ftla.  | 2e-38   | 1245686   | (U53181) F36D4.2 gene product [Caenorhabditis elegans]   | 0.19    |
| 1683                                  | U35032         | Human endogenous retrovirus clone c5.11, HERV-H multiply spliced subgenomic leader, protease and integrase region mRNA, partial cds | 1e-38   | 59977   | (Z14310) tripartite fusion transcript PLA2L [Human endogenous retrovirus]  | 1e-06   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1684   | D86974                                | Human mRNA for KIAA0220 gene, partial cds                                      | 1e-38   | 3337386  | (AC002544)<br>Unknown gene product splice form-2 [Homo sapiens] | 8e-11   |
| 1685   | M31013                                | Human nonmuscle myosin heavy chain (NMHC) mRNA, 3' end.                        | 1e-38   | 4115748  | (AB022023)<br>nonmuscle myosin heavy chain B                    | 2e-11   |
| 1686   | AF006087                              | Homo sapiens Arp2/3 protein complex subunit p20-Arc (ARC20) mRNA, complete cds | 4e-39   | <NONE>   | <NONE>  | <NONE>  |
| 1687   | X58374                                | D. melanogaster crn mRNA   | 4e-39   | 2655888  | (AL009171) 62D9.a [Drosophila melanogaster]                     | 4e-42   |
| 1688   | D85815                                | Human DNA for rhoHP1, complete cds   | 1e-39   | 134080   | GTP-BINDING PROTEIN TC10 ras-like protein [Homo sapiens]        | 3e-26   |
| 1689   | U49057                                | Rattus norvegicus CTD-binding SR-like protein rA9 mRNA, complete cds           | 4e-40   | 1438534  | (U49057) rA9 [Rattus norvegicus]                                | 5e-05   |
| 1690   | Y08999                                | H. sapiens mRNA for Sop2p-like protein   | 4e-40   | 3334339  | SOP2-LIKE PROTEIN   | 9e-08   |
| 1691   | AB002293                              | Human mRNA for KIAA0295 gene, partial cds                                      | 4e-40   | 2224531  | (AB002293)<br>KIAA0295 [Homo sapiens]                           | 1e-30   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1692   | AF086222                              | Homo sapiens full length insert cDNA clone ZC66E08                                       | 1e-40   | 2829669  | DOUBLE-STRANDED RNA-SPECIFIC EDITASE 1 (DSRNA ADENOSINE DEAMINASE) RNA EDITING ENZYME 1)<br>>gi 1707502 gnl PID e 254627 (X99227)<br>double-stranded RNA-specific editase [Homo sapiens]<br>editase 1 hRED1-L [Homo sapiens]<br>>gi 2039300 (U76421)<br>dsRNA adenosine deaminase DRADA2b [Homo sapiens] | 0.61    |
| 1693   | AF044127                              | Homo sapiens peroxisomal short-chain alcohol dehydrogenase (SCAD-SRL) mRNA, complete cds | 1e-40   | 4105190  | (AF044127)<br>peroxisomal short-chain alcohol dehydrogenase  | 2e-06   |
| 1694   | U36778                                | Mus musculus Sil mRNA, complete cds  | 1e-40   | 88608  | SIL protein - human<br>>gi 338088 (M74558)<br>SIL  | 6e-23   |
| 1695   | U36778                                | Mus musculus Sil mRNA, complete cds  | 1e-40   | 88608  | SIL protein - human<br>>gi 338088 (M74558)<br>SIL  | 6e-23   |
| 1696   | U36778                                | Mus musculus Sil mRNA, complete cds  | 1e-40   | 88608  | SIL protein - human<br>>gi 338088 (M74558)<br>SIL  | 5e-23   |
| 1697   | U36778                                | Mus musculus Sil mRNA, complete cds  | 1e-40   | 88608  | SIL protein - human<br>>gi 338088 (M74558)<br>SIL  | 5e-23   |
| 1698   | AB018285                              | Homo sapiens mRNA for KIAA0742 protein, partial cds                                      | 1e-40   | 3882205  | (AB018285)<br>KIAA0742 protein [Homo sapiens]  | 6e-31   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 1699   | X75927                                | M. musculus abc2 mRNA  | 1e-40   | 728773   | ATP-BINDING CASSETTE TRANSPORTER 1 ABC1-human >gi 495257 (X75926) abc1 [Mus musculus]   | 3e-37   |
| 1700   | AF038200                              | Homo sapiens clone 23954 mRNA sequence   | 5e-41   | 3211975  | (AF068195) putative glialblastoma cell differentiation-related protein [Homo sapiens]   | 5e-14   |
| 1701   | U20521                                | Human estrogen sulfotransferase (STE) gene, exon 8 and complete cds  | 4e-41   | <NONE>   | <NONE>  | <NONE>  |
| 1702   | AF026548                              | Homo sapiens branched chain alpha-ketoacid dehydrogenase kinase precursor, mRNA, nuclear gene encoding mitochondrial protein, complete cds | 2e-41   | 3182923  | [3-METHYL-2-OXOBUTANOATE DEHYDROGENASE (LIPOAMIDE)] KINASE PRECURSOR alpha-ketoacid dehydrogenase kinase precursor [Homo sapiens] | 2e-09   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1703                                  | Y07660         | M. tuberculosis accBC gene                           | 2e-41   | 465847  | HYPOTHETICAL 66.5 KD PROTEIN F02A9.5 IN CHROMOSOME III >gi 280542 pir  S28313 hypothetical protein F02A9.5 - Caenorhabditis elegans Genefinder; similar to Propionyl-CoA carboxylase beta chain; cDNA EST EMBL:M89018 comes from this gene; cDNA EST EMBL:D28069 comes from this gene; cDNA EST EMBL:D28068 comes from this gene; cDNA EST ... | 3e-38   |
| 1704                                  | AG001237       | Homo sapiens genomic DNA, 21q region, clone: 9H11N46 | 1e-41   | 106322  | Hypothetical protein (LIH 3' region) - human   | 5e-09   |
| 1705                                  | AB007934       | Homo sapiens mRNA for KIAA0465 protein, partial cds  | 1e-41   | 3413892   | (AB007934) KIAA0465 protein [Homo sapiens]   | 3e-12   |
| 1706                                  | AF055029       | Homo sapiens clone 24711 mRNA sequence               | 5e-42   | 3250681   | (AL024486) putative protein  | 2.2     |
| 1707                                  | Z49747         | O. cuniculus mRNA for phospholipase C                | 5e-42   | 130227  | 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE DELTA 1 (PLC-DELTA-1) (PHOSPHOLIPASE C-DELTA-1) (PLC-III) >gi 163538(M20638) phospholipase C-III [Box taurus]  | 5e-36   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1708   | M93651                                | Human set gene, complete cds.                            | 2e-42   | <NONE>   | <NONE>   | <NONE>  |
| 1709   | AJ236940                              | Sus scrofa mRNA for hypothetical protein (5'; clone 7C4) | 2e-42   | 2062403  | (U79010) delta 6 desaturase [Borago officinalis]                     | 8.5     |
| 1710   | J03634                                | Human erythroid differentiation protein mRNA             | 2e-42   | 1708436  | INHIBIN BETA A CHAIN PRECURSOR                                       | 2e-10   |
| 1711   | AJ223777                              | Mus musculus mRNA for striatin                           | 6e-43   | 2494917  | STRIATIN >gi 1495773 gn PID e254158                                  | 2e-32   |
| 1712   | AF016411                              | Homo sapiens potassium channel subunit KCNA3.1B          | 2e-43   | 2708514  | (AF016411) KCNA3.1B [Homo sapiens]                                   | 3e-13   |
| 1713   | AC001443                              | Homo sapiens (subclone 2_f10 from BAC2913                | 1e-43   | 111814   | Hypothetical protein 3 - rat >gi 56589                               | 2e-06   |
| 1714   | X82895                                | H. sapiens mRNA for DLG2                                 | 6e-44   | 2497511  | MAGUK P55 SUBFAMILY MEMBER 2 (MPP2 PROTEIN) (DISCS, LARGE HOMOLOG 2) | 6e-52   |
| 1715   | U17077                                | Human BENE mRNA, partial cds.                            | 3e-44   | 53912  | (X57960) ribosomal protein L7 [Mus musculus] >gi 55489               | 8e-30   |
| 1716   | AJ222700                              | Homo sapiens mRNA for TSC-22 protein                     | 2e-44   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION  | P VALUE |
| 1717   | J03634                                | Human erythroid differentiation protein mRNA   | 2e-44   | 124279   | INHIBIN BETA A CHAIN PRECURSOR PROTEIN) (EDF) >gi 87936 pir  B24248 inhibin beta-A chain precursor - human >gi 181947 (J03634) erythroid differentiation protein precursor [Homo sapiens] >gi 226850 prt  16082 60B inhibin betaA [Homo sapiens] | 0.73    |
| 1718   | AB014518                              | Homo sapiens mRNA for KIAA0618 protein, complete cds   | 7e-45   | 1911548  | (S80864) cytochrome c-like polypeptide sapiens]  | 1.6     |
| 1719   | X76808                                | H. sapiens genomic DNA clone d2  | 7e-45   | 868201   | (U29380) similar to adenylate cyclase [Caenorhabditis elegans]   | 2e-09   |
| 1720   | AB021288                              | Homo sapiens mRNA for beta 2-microglobulin, complete cds   | 2e-45   | 2465521  | (U95995) RNA-dependent RNA polymerase [Cryptosporidium parvum]   | 0.15    |
| 1721   | X63468                                | H. sapiens mRNA for transcription factor TFIIE alpha   | 8e-46   | <NONE>   | <NONE>   | <NONE>  |
| 1722   | AF019226                              | Homo sapiens D2-2 mRNA, 3'UTR  | 7e-46   | <NONE>   | <NONE>   | <NONE>  |
| 1723   | D31764                                | Human mRNA for KIAA0064 gene, complete cds   | 2e-46   | 3123050  | HYPOTHETICAL PROTEIN KIAA0064  | 1e-15   |
| 1724   | K02774                                | Human MHC class II HLA-DR-beta-psi (DW4/DR4) pseudogene, exons 3,4,5,6, clones cosII 3301 and cosII-801. | 1e-46   | 4185946  | (Y17834) gag protein [Human endogenous retrovirus K]   | 2e-14   |

| Nearest Neighbor (BlastN vs. Genbank) |            |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-SION | DESCRIPTION  | P VALUE | ACCES-SION  | DESCRIPTION  | P VALUE |
| 1725                                  | X92109     | H. sapiens hcglX gene                                      | 9e-47   | 2498185   | BRIDE OF SEVENLESS PROTEIN PRECURSOR<br>>gi 1079166 pir  A47550 bride of sevenless precursor - fruit fly (Drosophila virilis)<br>>gi 290216 virilis] | 1.4     |
| 1726                                  | X93334     | H. sapiens mitochondrial DNA, complete genome              | 8e-47   | 128753  | NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4<br>>gi 86696 pir  A00435 NADH dehydrogenase (ubiquinone)  | 4e-15   |
| 1727                                  | M85145     | Human tumor necrosis factor receptor, 3' flank.            | 3e-47   | <NONE>  | <NONE>   | <NONE>  |
| 1728                                  | X80240     | H. sapiens endogenous retrovirus HERV-KC4 DNA              | 3e-47   | 4185944   | (Y17833) env protein [Human endogenous retrovirus K]   | 7e-18   |
| 1729                                  | Z63594     | H. sapiens CpG DNA, clone 87f9, forward read cpg87t9.fila. | 1e-47   | 3322743   | (AE001222) T. pallidum predicted coding region TP0454  | 2.4     |
| 1730                                  | X62295     | R. rattus mRNA for vascular type-1 angiotensin II receptor | 4e-48   | 1209756   | (U43629 integral membrane protein [Beta vulgaris]  | 1e-07   |
| 1731                                  | M85145     | Human tumor necrosis factor receptor, 3' flank.            | 3e-48   | <NONE>  | <NONE>   | <NONE>  |
| 1732                                  | AB020712   | Homo sapiens mRNA for KIAA0905 protein, complete cds       | 4e-49   | 4240299   | (AB020712) KIAA0905 protein [Homo sapiens]   | 2e-20   |
| 1733                                  | AB020712   | Homo sapiens mRNA for KIAA0905 protein, complete cds       | 3e-49   | 4240299   | (AB020712) KIAA0905 protein [Homo sapiens]   | 2e-20   |
| 1734                                  | X62295     | R. rattus mRNA for vascular type-1 angiotensin II receptor | 1e-49   | 1209756   | (U43629) integral membrane protein [Beta vulgaris]   | 7e-12   |



| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1735                                  | AJ007509       | Homo sapiens mRNA for E1B-55kDa-associated protein                           | 1e-49   | 3319956   | (AJ007509) E1B-55kDa-associated protein  | 4e-24   |
| 1736                                  | X97303         | H. sapiens mRNA for Ptg-12 protein   | 1e-49   | 466044  | HYPOTHETICAL ZINC FINGER PROTEIN ZK686.4 IN CHROMOSOME III<br>>gi 630780 pir  S44909 ZK686.4 protein - Caenorhabditis elegans >gi 304346 (L17337) coded for by C. elegans cDNAs GenBank:M88869 and T01933; putative [Caenorhabditis elegans] | 8e-39   |
| 1737                                  | AF038404       | Homo sapiens homolog of Nedd5 (hNedd5) mRNA, complete cds                    | 4e-50   | <NONE>  | <NONE>   | <NONE>  |
| 1738                                  | L43618         | Homo sapiens polycystic kidney disease (PKD1) gene, exons 35-42              | 4e-50   | 903758  | (L43619) polycystic kidney disease 1 protein [Homo sapiens]  |         |
| 1739                                  | AF009424       | Homo sapiens clone 22 mRNA, alternative splice variant alpha-1, complete cds | 4e-50   | 2271473   | (AF009426) clone 22 [Homo sapiens]   |         |
| 1740                                  | L77040         | Homo sapiens (subclone 8_c11 from P1 H22) DNA sequence.                      | 2e-50   | 99758   | Monosaccharid transport protein STP4-Arabidopsis thaliana >gi 16524 (X66857) sugar transport protein [Arabidopsis thaliana]  | 6.4     |
| 1741                                  | L35657         | Homo sapiens (subclone H8 5_a10 from P1 35 H5 C8) DNA sequence.              | 2e-50   | 2072960   | (U93568) p40 [Homo sapiens]  | 2e-05   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION  | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1742   | U80745                                | Homo sapiens CTG7a mRNA, partial cds                   | 1e-50   | <NONE>   | <NONE>   | <NONE>  |
| 1743   | D84514                                | Bovine mRNA for p97, partial cds                       | 1e-50   | 3978527  | (AF103728) structural polyprotein [Sindbis virus]  | 9.9     |
| 1744   | M22960                                | Human protective protein mRNA, complete cds.           | 1e-50   | 131081   | LYSOSOMAL PROTECTIVE PROTEIN PRECURSOR (CATHEPSIN A) (CARBOXYPEPTIDASE C) human >gi 190283 (M22960) protective protein precursor                                       | 1e-12   |
| 1745   | X86018                                | H. sapiens mRNA for MUF1 protein                       | 1e-50   | 1082610  | Mufl protein - human >gi 762953 (X86018) muf1 [Homo sapiens]   | 1e-21   |
| 1746   | U03495                                | Human transcription factor LSF-ID mRNA, complete cds.  | 7e-51   | 2136296  | Transcription factor LSF - human >gi 476099  | 1e-21   |
| 1747   | AB015344                              | Homo sapiens HRIHFB2157 mRNA, partial cds              | 5e-51   | 3970874  | (AB015344) HRIHFB2157 [Homo sapiens]   | 2e-35   |
| 1748   | M93339                                | Human zinc finger protein mRNA.                        | 4e-51   | 3024110  | MYC-ASSOCIATED ZINC FINGER PROTEIN sapiens]  | 2e-06   |
| 1749   | U71363                                | Human zinc finger protein zfp6 (ZF6) mRNA, partial cds | 4e-51   | 2689441  | (AC003682) F18547_1 [Homo sapiens]   | 2e-11   |
| 1750   | X56932                                | H. sapiens mRNA for 23 kD highly basic protein         | 4e-51   | 730451   | 60S RIBOSOMAL PROTEIN L13A (23KD HIGHLY BASIC PROTEIN) >gi 345897 pir  S2953 9 basic protein, 23K - human >gi 23691 (X56932) 23 kD highly basic protein [Homo sapiens] | 1e-11   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|--|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1751                                  | Z79054         | H. sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA21E11 | 2e-51   | <NONE>  | <NONE>   | <NONE>  |
| 1752                                  | AF068245       | Homo sapiens BAF60b gene, partial sequence                       | 5e-52   | <NONE>  | <NONE>   | <NONE>  |
| 1753                                  | AJ236932       | Sus scrofa mRNA for hypothetical protein (5'; clone 4B8)         | 5e-52   | 400927  | RIBONUCLEOPROTEIN RB97D ribonucleoprotein [Drosophila melanogaster]  | 4.7     |
| 1754                                  | AF003693       | Mus musculus scaffold protein Pbp1 homolog mRNA, complete cds    | 6e-53   | 2197106   | (AF003693) scaffold protein Pbp1 homolog [Mus musculus]  | 2e-54   |
| 1755                                  | M27319         | Human calmodulin mRNA, complete cds.                             | 5e-53   | 115528  | CALMODULIN >gi 02408 pir  JC1309 calmodulin - Stylonychia lemnae (SGC5) >gi 161195                                       | 0.002   |
| 1756                                  | M74555         | Mouse house-keeping protein mRNA, complete cds.                  | 5e-53   | 284775  | House-keeping protein - mouse >gi 193871   | 5e-30   |
| 1757                                  | X92720         | H. sapiens mRNA for phosphoenolpyruvate carboxykinase            | 6e-54   | 2135915   | phosphoenolpyruvate carboxykinase (GTP) EC 4.1.1.32) precursor, mitochondrial - human carboxykinase (GTP) [Homo sapiens] | 6e-21   |
| 1758                                  | AF007872       | Homo sapiens torsinB (DQ1) mRNA, partial cds                     | 2e-54   | 2760121   | (AB002405) LAK-4p [Homo sapiens]   | 0.27    |
| 1759                                  | U49507         | Mus musculus B6CBA Lisch7 mRNA, partial cds.                     | 2e-54   | 1236083   | (U49507) Lisch7 [Mus musculus]   | 3e-27   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1760   | Z73360                                | Human DNA sequence from cosmid 92M18, BRCA2 gene region chromosome 13q12-13. | 1e-55   | 2370371  | (Y14657) hydrophobin [Pleurotus ostreatus] >gi 2982620 gnl PID e 1283986 (AJ225061) POH2 hydrophobin [Pleurotus ostreatus] | 2.0     |
| 1761   | U83702                                | Human cytochrome c oxidase subunit VIa gene, exon 3 and complete cds         | 8e-56   | 2982994  | (AE000682) hypothetical protein [Aquifex aeolicus]   | 7.0     |
| 1762   | Y12781                                | Homo sapiens mRNA for transducin (beta) like 1 protein.                      | 7e-56   | 3021409  | (Y12781) transducin (beta) like 1 protein [Homo sapiens]   | 7e-39   |
| 1763   | AB020673                              | Homo sapiens mRNA for KIAA0866 protein, complete cds                         | 8e-57   | 2104553  | (AF001548) Myosin heavy chain (MHY11) (5' partial) [Homo sapiens]  | 4e-04   |
| 1764   | AJ236932                              | Sus scrofa mRNA for hypothetical protein (5'; clone 4B8)                     | 3e-57   | 400927   | RIBONUCLEOPROTEIN RB97D ribonucleoprotein [Drosophila melanogaster]  | 4.7     |
| 1765   | L06900                                | Human dystrophin gene, intron 1 containing pseudo exon.                      | 1e-58   | 4185129  | (AC005724) unknown protein [Arabidopsis thaliana] thaliana]  | 7.0     |
| 1766   | X93334                                | H. sapiens mitochondrial DNA, complete genome                                | 9e-59   | 1492050  | (U60315) MC107L [Mollusum contagiosum virus subtype 1]   | 0.17    |
| 1767   | AF064856                              | Rattus sp. 7acomp protein mRNA, complete cds                                 | 3e-59   | 3169626  | (AF064856) 7acomp protein [Rattus sp.]   | 2e-31   |
| 1768   | AF081484                              | Homo sapiens alpha-tubulin isoform 1 mRNA, complete cds                      | 2e-59   | 32015  | (X06956) alpha-tubulin [Homo sapiens]  | 4e-22   |
| 1769   | X71427                                | Homo sapiens mRNA for FUS-CHOP protein fusion                                | 1e-60   | 746557   | (U23523) histidine-rich [Caenorhabditis elegans]   | 0.45    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION   | P VALUE |
| 1770   | AF013988                              | Homo sapiens serine protein mRNA, complete cds  | 1e-60   | 2564316  | (AB006622) No similarities to any reported proteins [Homo sapiens]  | 0.26    |
| 1771   | U25691                                | Mus musculus lymphocyte specific helicase mRNA, complete cds  | 7e-61   | 2137490  | lymphocyte specific helicase - mouse musculus]  | 3e-25   |
| 1772   | X93334                                | H. sapiens mitochondrial DNA, complete genome   | 4e-61   | 70656  | ubiquitin / ribosomal protein S27a - human extension protein, HUBCEP80 [human, Peptide, 156 aa] ubiquitin extention protein [Cavia porcellus] | 9e-08   |
| 1773   | D38255                                | Homo sapiens mRNA for CAB1, complete cds  | 4e-61   | 2135214  | Gene MLN 64 protein - human   | 4e-23   |
| 1774   | U25691                                | Mus musculus lymphocyte specific helicase mRNA, complete cds  | 8e-62   | 2137490  | lymphocyte specific helicase-mouse musculus]  | 8e-26   |
| 1775   | M21731                                | Human lipocortin-V mRNA, complete cds.  | 6e-62   | 3212603  | Human Annexin V With Proline Substitution By Thioproline  | 2e-20   |
| 1776   | AF021936                              | Rattus norvegicus myotonic dystrophy kinase-related Cdc42-binding kinase MRCK-beta (MRCK-beta) mRNA, complete cds | 2e-62   | 2736153  | (AF021936) myotonic dystrophy kinase-related Cdc42-binding kinase MRCK-beta [Rattus norvegicus]   | 3e-27   |
| 1777   | Y12059                                | H. sapiens HUNKI mRNA   | 1e-62   | 3184498  | (AC004798) R31546_1 [Homo sapiens]  | 3e-09   |
| 1778   | L37368                                | Human (clone E5.1) RNA-binding protein mRNA, complete cds.  | 6e-63   | 477578   | sialidase - Actinomyces viscosus >gi 41852  | 7.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION  | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1779   | M27877                                | Figure 2. Nucleotide and translated protein sequences of HPF1, -2, and -9. | 5e-63   | 1731443  | ZINC FINGER PROTEIN 83 (ZINC FINGER PROTEIN HPF1)<br>>gi 106023 pir  A32891 finger protein 1, placental - human  | 3e-33   |
| 1780   | AF095448                              | Homo sapiens putative G protein-coupled receptor                           | 2e-63   | 3116131  | (AL023288)<br>hypothetical protein   | 4.6     |
| 1781   | L19437                                | Human transaldolase mRNA containing transposable element, complete cds     | 2e-63   | 1553119  | (U63159)<br>transaldolase [Mus musculus]   | 4e-18   |
| 1782   | L41351                                | Homo sapiens prostasin mRNA, complete cds                                  | 1e-63   | 2833277  | PROSTASIN PRECURSOR<br>precursor - human<br>>gi 862305 (L41351)<br>prostasin [Homo sapiens] >gi 1143194 (U33446) prostasin [Homo sapiens]  | 6e-14   |
| 1783   | AF053470                              | Homo sapiens 10kD protein (BC10) mRNA, complete cds                        | 6e-64   | 482237   | Hypothetical protein K03H1.9 - Caenorhabditis elegans  | 0.029   |
| 1784   | D37791                                | Mouse mRNA for bcta-1,4-galactosyltransferase                              | 6e-64   | 3880102  | (Z93390) similar to FYVE zinc finger; cDNA EST yk265b4.5 comes from this gene; cDNA EST yk359g9.5 comes from this gene; cDNA EST yk319c2.5 comes from this gene [Caenorhabditis elegans] zinc finger; cDNA EST yk265b4.5 comes from this gene; cDNA EST yk359g9.5 comes from this gene; cDNA EST yk319c2.5 comes from this gene [Caenorhabditis elegans] | 3e-16   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1785   | AF015770                              | Mus musculus radical fringe (radical-fringe) mRNA, complete cds   | 6e-64   | 2204355  | (U94350) radical fringe precursor [Mus musculus]   | 1e-36   |
| 1786   | Z79054                                | H. sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA21E11  | 2e-64   | <NONE>   | <NONE>   | <NONE>  |
| 1787   | M83094                                | Homo sapiens cytosolic selenium-dependent glutathione peroxidase gene, complete cds, and rhoh 12 gene, 3'end. | 1e-64   | 2447063  | (U42580) A565R [Paramecium bursaria Chlorella virus 1]   | 8.8     |
| 1788   | Y10211                                | H. sapiens LAG-3 gene, promoter region  | 7e-65   | 1944540  | (X14112) tegument protein [human herpesvirus 1]  | 2.3     |
| 1789   | M19045                                | Human lysozyme mRNA, complete cds.  | 2e-65   | <NONE>   | <NONE>   | <NONE>  |
| 1790   | U01882                                | Homo sapiens SS-A/Ro autoantigen 52 kda component gene, complete cds  | 2e-65   | 585401   | LIPASE MODULATOR PRECURSOR (LIPASE HELPER PROTEIN)<br>>gi 480045 pir  S3624 9 lipB protein - Pseudomonas glumae<br>>gi 49207 (X70354) helper protein | 4.2     |
| 1791   | AF069517                              | Homo sapiens RNA binding protein DEF-3 mRNA, complete cds   | 2e-65   | 3212101  | (AF069517) RNA binding protein DEF-3 [Homo sapiens]  | 1e-25   |
| 1792   | AF004715                              | Homo sapiens jerky gene product homolog mRNA, complete cds  | 2e-65   | 2314829  | (AF004715) jerky gene product homolog [Homo sapiens]   | 2e-45   |
| 1793   | X59652                                | C. longicaudatus hppt mRNA for hypoxanthine   | 3e-66   | 631625   | hypoxanthine (guanine) phosphoribosyltransferase - long-tailed hamster phosphoribosyltransferase [Cricetulus longicaudatus]                          | 6e-54   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1794                                  | U934350        | Mus musculus radical fringe precursor mRNA, complete cds                                      | 3e-67   | 2204355   | (U94350) radical fringe precursor [Mus musculus]   | 2e-33   |
| 1795                                  | AF015811       | Mus musculus putative lysophosphatidic acid acyltransferase mRNA, complete cds                | 3e-68   | 2317725   | (AF015811) putative lysophosphatidic acid acyltransferase [Mus musculus]   | 7e-51   |
| 1796                                  | J03137         | Cow phosphoinositide-specific phospholipase C   | 3e-69   | 226908  | phospholipase C 154 [Bos taurus]   | 3e-25   |
| 1797                                  | AF044574       | Rattus norvegicus putative peroxisomal 2,4-dienoyl-CoA reductase (DCR-AKL) mRNA, complete cds | 1e-69   | 4105269   | (AF044574) putative peroxisomal 2,4-dienoyl-CoA reductase [Rattus norvegicus]  | 2e-33   |
| 1798                                  | AF015811       | Mus musculus putative lysophosphatidic acid acyltransferase mRNA, complete cds                | 4e-70   | 2317725   | (AF015811) putative lysophosphatidic acid acyltransferase [Mus musculus]   | 3e-19   |
| 1799                                  | X65157         | M. musculus mRNA for desmoyokin, partial  | 5e-74   | 109781  | desmoyokin - mouse (fragment) >gi 50675  | 9e-37   |
| 1800                                  | Z97207         | Mus musculus mRNA for B-IND1 protein  | 2e-74   | 2231019   | (Z97207) B-IND1 protein [Mus musculus]   | 6e-21   |
| 1801                                  | U27196         | Gallus gallus zinc finger protein (Fzf-1) mRNA, complete cds.                                 | 6e-75   | 984814  | (U27196) zinc finger protein [Gallus gallus]   | 2e-44   |
| 1802                                  | Y15054         | Rattus norvegicus mRNA for 70 kDa tumor specific antigen, partial                             | 3e-77   | 3123027   | 70 KD WD-REPEAT TUMOR-SPECIFIC ANTIGEN >gi 2505957 gnl PID e353992 (Y15054) 70 kD tumor-specific antigen [Rattus norvegicus] | 4e-42   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES-SION                            | DESCRIPTION   | P VALUE | ACCES-SION   | DESCRIPTION  | P VALUE |
| 1803   | X65157                                | M. musculus mRNA for desmoyokin, partial  | 3e-79   | 109781   | desmoyokin - mouse (fragment) >gi 50675  | 9e-33   |
| 1804   | U50736                                | Rattus norvegicus cardiac adriamycin responsive protein mRNA, complete cds                                      | 2e-84   | 1362781  | Cytokine inducible nuclear protein C193 - human >gi 793841 (X83703) nuclear protein [Homo sapiens] | 7e-30   |
| 1805   | AF072865                              | Rattus norvegicus thioredoxin reductase (TrxR2) mRNA, nuclear gene encoding mitochondrial protein, complete cds | 2e-84   | 3757888  | (AF072865) thioredoxin reductase [Rattus norvegicus]   | 6e-43   |
| 1806   | AF044574                              | Rattus norvegicus putative peroxisomal 2,4-dienoyl-CoA reductase (DCR-AKL) mRNA, complete cds                   | 6e-85   | 4105269  | (AF044574) putative peroxisomal 2,4-dienoyl-CoA reductase [Rattus norvegicus]                      | 1e-41   |
| 1807   | U19181                                | Rattus norvegicus Rabin3 mRNA, complete cds.  | 2e-87   | 624225   | (U19181) Rabin3 [Rattus norvegicus]  | 2e-41   |
| 1808   | U40342                                | Mus musculus ninein mRNA, complete cds.   | 1e-91   | 1113865  | (U40342) ninein [Mus musculus]   | 2e-36   |
| 1809   | X67877                                | R. norvegicus mRNA for cytosolic resiniferatoxin-binding protein  | 4e-92   | 136077   | TROPOMYOSIN BETA 3, FIBROBLAST chicken >gi 515694 (M23082) tropomyosin [Gallus gallus]             | 0.56    |
| 1810   | AF044574                              | Rattus norvegicus putative peroxisomal 2,4-dienoyl-CoA reductase (DCR-AKL) mRNA, complete cds                   | 5e-93   | 4105269  | (AF044574) putative peroxisomal 2,4-dienoyl-CoA reductase [Rattus norvegicus]                      | 1e-50   |
| 1811   | AF035527                              | Mus musculus EHF (Ehf) mRNA, complete cds   | 2e-95   | 3138930  | (AF035527) EHF [Mus musculus]  | 2e-47   |

| Nearest Neighbor (BlastN vs. Genbank) |                |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|--|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION  | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1812                                  | AB016930       | Cricetus griseus mRNA for Phosphatidylglycerophosphate synthase, complete cds  | 6e-96   | 4159682   | (AB016930) Phosphatidylglycerophosphate synthase [Cricetus griseus]   | 7e-41   |
| 1813                                  | AB005549       | Rattus norvegicus mRNA for atypical PKC specific binding protein, complete cds   | 7e-97   | 3868778   | (AB005549) atypical PKC specific binding protein [Rattus norvegicus]  | 3e-41   |
| 1814                                  | X90849         | G. gallus PB1 gene   | 2e-97   | 2134381   | polybromo 1 protein - chicken chicken >gi 951231 (X90849) polybromo 1 protein [Gallus gallus]   | 1e-34   |
| 1815                                  | S79873         | h-lamp-2=lysosome-associated membrane protein-2 protein-2b (LAMP2) mRNA, alternatively spliced form h-lamp-2b, complete cds. | 3e-98   | <NONE>  | <NONE>  | <NONE>  |
| 1816                                  | U67203         | Mus musculus ACF7 neural isoform 1 (mACF7) mRNA, partial cds   | 2e-98   | 1675224   | (U67204) ACF7 neural isoform 2 [Mus musculus]   | 9e-39   |
| 1817                                  | L14684         | Rattus norvegicus nuclear-encoded mitochondrial elongation factor G mRNA, complete cds.                                      | e-100   | 585084  | ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G) >gi 543383 pir  S40780 translation elongation factor G, mitochondrial - rat >gi 310102 | 2e-30   |
| 1818                                  | X84692         | M. musculus Spnr mRNA for RNA binding protein  | e-133   | 1363238   | Spermatid perinuclear RNA-binding protein Spnr - mouse >gi 673454 (X84692) spermatid perinuclear RNA binding protein [Mus musculus]         | 5e-35   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|----------------|---|---------|---|---|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION   | P VALUE |
| 1819                                  | U50736         | Rattus norvegicus<br>cardiac adriamycin<br>responsive protein<br>mRNA, complete cds | e-113   | 1362781   | Cytokine inducible<br>nuclear protein C193 -<br>human >gi 793841<br>(X83703) nuclear<br>protein [Homo<br>sapiens] | 2e-36   |
| 1820                                  | S66855         | HoxB9=Hox-2.5<br>[mice, embryos,<br>mRNA Partial, 786 nt]                           | e-107   | 1708355   | HOMEBOX<br>PROTEIN HOX-B9<br>(HOX-2.5)  | 8e-37   |
| 1821                                  | S66855         | HoxB9=Hox-2.5<br>[mice, embryos,<br>mRNA Partial, 786 nt]                           | e-108   | 1708355   | HOMEBOX<br>PROTEIN HOX-B9<br>(HOX-2.5)  | 4e-37   |
| 1822                                  | U92072         | Rattus norvegicus m-<br>tomosyn mRNA,<br>complete cds                               | e-102   | 3790389   | (U92072) m-tomosyn<br>[Rattus norvegicus]   | 2e-38   |
| 1823                                  | D17577         | Mouse mRNA for<br>kinesin-like protein<br>(Kif1b), complete cds                     | e-129   | 2497524   | KINESIN-LIKE<br>PROTEIN KIF1B<br>mouse<br>>gi 407339 gnl PID d1<br>005029 (D17577)<br>Kif1b [Mus musculus]        | 2e-39   |
| 1824                                  | AF062484       | Mus musculus SDP8<br>mRNA, complete cds   | e-122   | 3126981   | (AF062484) SDP8<br>[Mus musculus]   | 5e-40   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION   | P VALUE |
| 1825   | X73683                                | R. norvegicus mRNA for histone H3.3   | e-109   | 122075   | Histone H3.3 (H3.B) (H3.3Q) histone H3.3 - fruit fly (Drosophila melanogaster) histone H3.3B - chicken >gi 2119023 pir  S61218 histone H3.3 - fruit fly (Drosophila hydei) 1-136 [Oryctolagus cuniculus] >gi 8046(X53822) Histone H3.3Q gene product [Drosophila melanogaster] >>gi 51998 gallus] >gi 161190 (M17876) histone H3 [Spisula solidissima] >gi 211853 (M11393) histone 3.3 [Gallus gallus] >gi 306848 (M11354) H3.3 histone [Homo sapiens] melanogaster] >gi 963031 (X81205) histone H3.3 H3.3A variant [Drosophila melanogaster] musculus] | 2e-40   |
| 1826   | U67203                                | Mus musculus ACF7 neural isoform 1 (mACF7) mRNA, partial cds                    | e-102   | 1675224  | (U67204) ACF7 neural isoform 2 [Mus musculus]   | 2e-40   |
| 1827   | D17577                                | Mouse mRNA for kinesin-like protein (Kiflb), complete cds                       | e-131   | 2497524  | KINESIN-LIKE PROTEIN KIF1B mouse >gi 407399 gnl PID d1005029 (D17577) Kiflb [Mus musculus]  | 7e-42   |
| 1828   | AB016930                              | Cricetulus griseus mRNA for Phosphatidylglycerophosphate synthase, complete cds | e-131   | 4159682  | (AB016930) Phosphatidylglycerophosphate synthase [Cricetulus griseus]   | 3e-43   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCES-<br>SION                        | DESCRIPTION  | P VALUE | ACCES-<br>SION                                       | DESCRIPTION   | P VALUE |
| 1829   | U09874                                | Mus musculus SKD3 mRNA, complete cds.  | e-122   | 2493735  | SKD3 PROTEIN SKD3 [Mus musculus]  | 7e-48   |
| 1830   | X99145                                | C. familiaris mRNA for C3VS protein  | e-110   | 1429314  | (X99145) overexpressed in thyroid tissue after TSH stimulation [Canis familiaris]   | 2e-49   |
| 1831   | X99836                                | P. walti mRNA for rnp associated protein 55  | e-106   | 4200286  | (X99836) rap55 [Pleurodeles waltl]  | 2e-50   |
| 1832   | AF077003                              | Mus musculus SH3 domain-containing adapter protein mRNA, complete cds                          | e-121   | 3550240  | (AF077003) SH3 domain-containing adapter protein; CD2AP   | 3e-51   |
| 1833   | AF060246                              | Mus musculus strain C57BL/6 zinc finger protein 106 (Zfp 106) mRNA, H3a-a allele, complete cds | e-118   | 3372657  | (AF060246) zinc finger protein 106 [Mus musculus]   | 1e-52   |
| 1834   | Z14030                                | R. norvegicus mRNA for TRAP-complex gamma subunit.   | e-120   | 1174453  | TRANSLOCON-ASSOCIATED PROTEIN GAMMA SUBUNIT (TRAP-GAMMA) (SIGNAL SEQUENCE RECEPTOR GAMMA SUBUNIT) (SSR-GAMMA) >gi 423185 pir  S3329 4 translocon-associated protein gamma chain - rat norvegicus] | 7e-54   |
| 1835   | AF077003                              | Mus musculus SH3 domain-containing adapter protein mRNA, complete cds                          | e-132   | 3550240  | (AF077003) SH3 domain-containing adapter protein; CD2AP   | 5e-54   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCES- SION                           | DESCRIPTION   | P VALUE | ACCES- SION  | DESCRIPTION  | P VALUE |
| 1836   | L20427                                | Rattus norvegicus dihydroxypolyprenylb enzoate methyltransferase mRNA, complete cds                             | e-116   | 457372   | (L20427) dihydroxypolyprenylb enzoate methyltransferase dihydroxypolyprenylb enzoate methyltransferase [Rattus norvegicus] | 4e-56   |
| 1837   | X80169                                | M. musculus mRNA for 200 kD protein   | e-122   | 1717793  | PROTEIN TSG24 (MEIOTIC CHECK POINT REGULATOR) >gi 083553 pir  A55117 tsg24   | 2e-56   |
| 1838   | AF080568                              | Rattus norvegicus CTP:phosphoethanol amine cytidyltransferase mRNA, complete cds                                | e-119   | 3396102  | (AF080568) CTP:phosphoethanolamine cytidyltransferase  | 6e-58   |
| 1839   | X99145                                | C. familiaris mRNA for C3VS protein   | e-121   | 1429314  | (X99145) overexpressed in thyroid tissue after TSH stimulation [Canis familiaris]  | 2e-58   |
| 1840   | AF019075                              | Pan troglodytes breast and ovarian cancer susceptibility (BRCA1) gene, partial cds                              | e-145   | 2218154  | (AF005068) breast and ovarian cancer susceptibility protein splice variant [Homo sapiens]                                  | 1e-58   |
| 1841   | U55042                                | Box taurus myosin X, complete cds   | e-122   | 1755049  | (U55042) myosin X [Bos taurus]   | 1e-61   |
| 1842   | AJ007780                              | Mus musculus mRNA for poly (ADP-ribose) polymerase-2  | e-119   | 3283975  | (AF072521) poly-(ADPriboseyl)-transferase homolog PARP   | 4e-62   |
| 1843   | AF072865                              | Rattus norvegicus thioredoxin reductase (TrxR2) mRNA, nuclear gene encoding mitochondrial protein, complete cds | e-105   | 3757888  | (AF072865) thioredoxin reductase [Rattus norvegicus]   | 3e-62   |

| Nearest Neighbor (BlastN vs. Genbank) |                |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|----------------|---|---------|---|--|---------|
| SEQ ID                                | ACCES-<br>SION | DESCRIPTION   | P VALUE | ACCES-<br>SION  | DESCRIPTION  | P VALUE |
| 1844                                  | U55042         | Bos taurus myosin X,<br>complete cds                    | e-121   | 1755049   | (U55042) myosin X<br>[Bos taurus]  | 1e-62   |
| 1845                                  | X61506         | Mouse E46 mRNA for<br>E46 protein                       | e-139   | 114909  | BRAIN PROTEIN<br>E46   | 9e-67   |
| 1846                                  | D90335         | Bovine mRNA for<br>GTP-binding protein<br>alpha-subunit | e-148   | 585174  | GUANINE<br>NUCLEOTIDE-<br>BINDING PROTEIN,<br>ALPHA-14<br>SUBUNIT (GL1)<br>>gi 108711 pir  A4089<br>1 GTP-binding protein<br>GL1 alpha chain -<br>bovine protein, alpha-<br>subunit [Bos taurus] | 2e-69   |
| 1847                                  | U49507         | Mus musculus<br>B6CBA Lisch7<br>mRNA, partial cds.      | e-140   | 2121326   | (AC002128) Lisch7<br>[Homo sapiens]  | 2e-74   |

TABLE 4

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|-------------|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 2         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 3         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 4         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 5         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 6         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 7         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 8         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 9         | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 10        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 11        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 12        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 13        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 14        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 15        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 16        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 17        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 18        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 19        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 20        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 21        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 22        | <NONE>                                | <NONE>      | <NONE>  | <NONE>  | <NONE>   | <NONE>  |
| 23        | <NONE>                                | <NONE>      | <NONE>  | 1079469   | tMDC I protein - crab-eating macaque                                     | 9.3     |
| 24        | <NONE>                                | <NONE>      | <NONE>  | 3043656   | (AB011138)<br>KIAA0566 protein<br>[Homo sapiens]                         | 9.3     |
| 25        | <NONE>                                | <NONE>      | <NONE>  | 112175  | potassium channel<br>protein RK5 - rat<br>protein [Rattus<br>norvegicus] | 8.6     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|-------------|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 26     | <NONE>                                | <NONE>      | <NONE>  | 3769624  | (AF091565) olfactory receptor [Rattus norvegicus]  | 7.2     |
| 27     | <NONE>                                | <NONE>      | <NONE>  | 3876443  | (Z81517) F28B1.6 [Caenorhabditis elegans]  | 7.1     |
| 28     | <NONE>                                | <NONE>      | <NONE>  | 2224464  | (AB001684) ORF249 [Chlorella vulgaris]   | 6.9     |
| 29     | <NONE>                                | <NONE>      | <NONE>  | 1519707  | (U67940) ORFveg106; random cDNA sequence [Dictyostelium discoideum]  | 6.7     |
| 30     | <NONE>                                | <NONE>      | <NONE>  | 227491   | protein kinase C II [Xenopus laevis]   | 6.7     |
| 31     | <NONE>                                | <NONE>      | <NONE>  | 630575   | C50C3.4 protein - Caenorhabditis elegans   | 6.0     |
| 32     | <NONE>                                | <NONE>      | <NONE>  | 137290   | 35 KD PROTEIN IN RNA2 clover necrotic mosaic virus >gi 61466 (X08021) ORF for 35 kDa polypeptide (AA 1-317) [Red clover necrotic mosaic virus] | 6.0     |
| 33     | <NONE>                                | <NONE>      | <NONE>  | 30041  | (X16711) pid:g30041 [Homo sapiens]   | 5.9     |
| 34     | <NONE>                                | <NONE>      | <NONE>  | 2493585  | CELL DIVISION PROTEIN FTSW   | 5.7     |
| 35     | <NONE>                                | <NONE>      | <NONE>  | 1001450  | (D63999) hypothetical protein  | 5.7     |
| 36     | <NONE>                                | <NONE>      | <NONE>  | 3182918  | NITROGEN REGULATORY PROTEIN AREA   | 5.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|-------------|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 37     | <NONE>                                | <NONE>      | <NONE>  | 140011   | MITOCHONDRIAL RIBOSOMAL PROTEIN S5<br>Emericella nidulans mitochondrion (SGC3)<br>>gi 12709 nidulans]<br>>gi 472822 (J01390)<br>unknown protein                                       | 4.3     |
| 38     | <NONE>                                | <NONE>      | <NONE>  | 3979943  | (AL034393) predicted using Genefinder; similar to WD domain, G-beta repeat; cDNA EST yk362f7.5 comes from this gene; cDNA EST yk362f7.3 comes from this gene [Caenorhabditis elegans] | 4.0     |
| 39     | <NONE>                                | <NONE>      | <NONE>  | 950203   | (U31329) polyketide synthase [Aspergillus terreus]  | 3.3     |
| 40     | <NONE>                                | <NONE>      | <NONE>  | 3560232  | (AL031530) hypothetical zinc finger protein [Schizosaccharomyces pombe]   | 3.0     |
| 41     | <NONE>                                | <NONE>      | <NONE>  | 730071   | AXONEME-ASSOCIATED PROTEIN MST101(1) product [Drosophila hydei]   | 2.6     |
| 42     | <NONE>                                | <NONE>      | <NONE>  | 2506641  | HYPOTHETICAL 21.7 KD PROTEIN IN INTE-PIN INTERGENIC REGION >gi 1787402 (AE000214) orf, hypothetical protein [Escherichia coli]  | 2.5     |
| 43     | <NONE>                                | <NONE>      | <NONE>  | 3511232  | (AF071556) anthranilate dioxygenase large subunit   | 2.4     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|-------------|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 44     | <NONE>                                | <NONE>      | <NONE>  | 1150900  | (U43139) envelope glycoprotein gp120 [Human immunodeficiency virus type 1]   | 1.9     |
| 45     | <NONE>                                | <NONE>      | <NONE>  | 3876099  | (Z75536) similar to dynein heavy chain; cDNA EST EMBL:D27549 comes from this gene; cDNA EST EMBL:D34859 comes from this gene [Caenorhabditis elegans]  | 1.4     |
| 46     | <NONE>                                | <NONE>      | <NONE>  | 3881150  | (AL032647) predicted using Genefinder  | 1.4     |
| 47     | <NONE>                                | <NONE>      | <NONE>  | 132200   | COLANIC ACID CAPSULAR BIOSYNTHESIS ACTIVATION PROTEIN A >gi 95605 pir  S17701 rcsA protein   | 1.1     |
| 48     | <NONE>                                | <NONE>      | <NONE>  | 2204286  | (U61380) germination protein [Bacillus megaterium]   | 1.0     |
| 49     | <NONE>                                | <NONE>      | <NONE>  | 1723955  | HYPOTHETICAL 11.4 KD PROTEIN IN FOX1-KEX1 INTERGENIC REGION >gi 2132566 pir  S6422 2 probable membrane protein YGL204c - yeast (Saccharomyces cerevisiae) >gi 1322838 gnl PID e2 43803 (Z72726) ORF YGL204c [Saccharomyces cerevisiae] | 0.84    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|-------------|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 50     | <NONE>                                | <NONE>      | <NONE>  | 3201564  | (AJ006514) prolipoprotein diacylglycerol transferase [Vibrio cholerae]   | 0.31    |
| 51     | <NONE>                                | <NONE>      | <NONE>  | 2808721  | (AL021428) hypothetical protein Rv0064   | 0.27    |
| 52     | <NONE>                                | <NONE>      | <NONE>  | 602434   | (U17986) GABA/noradrenaline transporter [Homo sapiens]   | 0.13    |
| 53     | <NONE>                                | <NONE>      | <NONE>  | 3347955  | (AF076184) cytosolic sorting protein PACS-1b [Rattus norvegicus]   | 0.12    |
| 54     | <NONE>                                | <NONE>      | <NONE>  | 1255887  | (U53344) coded for by C. elegans cDNA yk92b4.5; coded for by C. elegans cDNA yk73a1.5; coded for by C. elegans cDNA yk102e9.5; coded for by C. elegans cDNA yk71c8.5; coded for by C. elegans cDNA yk66d11.5; coded for by C. elegans cDNA yk66c3... | 0.074   |
| 55     | <NONE>                                | <NONE>      | <NONE>  | 103076   | Bkm-like sex-determining region hypothetical protein CS314 - fruit fly (Drosophila melanogaster)   | 0.003   |
| 56     | <NONE>                                | <NONE>      | <NONE>  | 107560   | Ras inhibitor (clone JC265) - human sapiens]   | 0.002   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |             |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|-------------|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 57     | <NONE>                                | <NONE>      | <NONE>  | 103076   | Bkm-like sex-determining region hypothetical protein CS314 - fruit fly ( <i>Drosophila melanogaster</i> )   | 2e-04   |
| 58     | <NONE>                                | <NONE>      | <NONE>  | 2702370  | (AF038604) contains similarity to <i>Drosophila</i> ovarian tumor locus protein (GB:X13693) [ <i>Caenorhabditis elegans</i> ]   | 6e-05   |
| 59     | <NONE>                                | <NONE>      | <NONE>  | 3859713  | (AL033501) phox domain protein [ <i>Candida albicans</i> ]  | 3e-05   |
| 60     | <NONE>                                | <NONE>      | <NONE>  | 2088839  | (AF003386) F59E12.5 gene product [ <i>Caenorhabditis elegans</i> ]  | 2e-08   |
| 61     | <NONE>                                | <NONE>      | <NONE>  | 121059   | GC-RICH SEQUENCE DNA-BINDING FACTOR GCF - human >gi 179412 (M29204) DNA-binding factor [ <i>Homo sapiens</i> ]  | 4e-09   |
| 62     | <NONE>                                | <NONE>      | <NONE>  | 3875246  | (Z81490) similar to WD domain, G-beta repeats (2 domains); cDNA EST EMBL:T00482 comes from this gene; cDNA EST EMBL:T00923 comes from this gene; cDNA EST yk449d4.3 comes from this gene; cDNA EST yk449d4.5 comes from this gen... | 9e-24   |
| 63     | <NONE>                                | <NONE>      | <NONE>  | 1465834  | (U64857) No definition line found [ <i>Caenorhabditis elegans</i> ]   | 9e-28   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 64     | <NONE>                                | <NONE>   | <NONE>  | 3327136  | (AB014561) KIAA0661 protein [Homo sapiens]  | 1e-29   |
| 65     | <NONE>                                | <NONE>   | <NONE>  | 3880433  | (Z66521) similar to mitochondrial RNA splicing MSR4 like protein; cDNA EST EMBL:C09217 comes from this gene [Caenorhabditis elegans]  | 8e-31   |
| 66     | D42133                                | Rat annexin V gene, exon7 and exon8  | 5.0     | <NONE>   | <NONE>  | <NONE>  |
| 67     | L35679                                | Homo sapiens (subclone H8 2_d11 from P1 35 H5 C8) DNA sequence.                        | 5.0     | 1086902  | (U41278) coded for by C. elegans cDNA yk79g8.5; coded for by C. elegans cDNA cm10c8; coded for by C. elegans cDNA yk79g8.3; similar to leucine-rich repeats found in many proteins [Caenorhabditis elegans] | 6.6     |
| 68     | U90184                                | HIV-1 strain BX220 from USA, envelope glycoprotein C2V3 region (env) gene, partial cds | 5.0     | 1297070  | (Z71986) convicilin precursor [Vicia narbonensis]   | 6.6     |
| 69     | U61465                                | Human myosin VIIa (MYO7A) gene, 5' exon 37   | 5.0     | 2313225  | (AE000535) L-lactate permease (lctP) [Helicobacter pylori 26695]  | 5.0     |
| 70     | AF013717                              | Homo sapiens periplakin (PPL) mRNA, partial cds  | 5.0     | 3719238  | (AF064869) brain-enriched guanylate kinase-associated protein 2; BEGA2 [Rattus norvegicus]  | 3.8     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 71                                    | X58245    | Soybean mRNA for HMG-1 like protein  | 5.0     | 2995363   | (AL022245) biotin synthase   | 0.99    |
| 72                                    | AF102425  | Fraseria paniculata tRNA-Leu (trnL) gene, intron, chloroplast sequence       | 4.9     | 3522958   | (AC004411) putative pectinesterase [Arabidopsis thaliana]  | 6.4     |
| 73                                    | X82817    | H.sapiens PTP1C/HCP-variant gene   | 4.9     | 3875514   | (Z81494) cDNA EST EMBL:D27474 comes from this gene; cDNA EST EMBL:D27473 comes from this gene; cDNA EST EMBL:T00471 comes from this gene; cDNA EST EMBL:D34192 comes from this gene; cDNA EST EMBL:D37241 comes from this gene; ...  | 2.8     |
| 74                                    | U04827    | Mus musculus brain fatty acid-binding protein                                | 4.9     | 3676132   | (AL031765) I-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=31.96; 1-evidence_end; 2-evidence=predicted by match; 2-match_accession=SPT REMBL:Q93319; 2-match_description=HYPOTHETICAL PROTEIN C33A11.2.;... | 2e-09   |
| 75                                    | AF038859  | Neospora hughesi strain NE1 internal transcribed spacer 1, complete sequence | 4.8     | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 76     | Y08222                                | M.musculus MFH-1 gene   | 4.8     | <NONE>   | <NONE>   | <NONE>  |
| 77     | AJ224475                              | Borrelia burgdorferi left chromosomal subtelomeric region (pfpB gene)                 | 4.8     | 4218141  | (AJ236702) HMR1 protein [Antirrhinum majus]                        | 8.3     |
| 78     | U02486                                | Mus musculus LAF putative membrane protein (KRAG) gene, exon 3 and complete cds       | 4.8     | 3258103  | (AP000006) 367aa long hypothetical protein [Pyrococcus horikoshii] | 2.7     |
| 79     | AB000280                              | Rat mRNA for peptide/histidine transporter, complete cds                              | 4.8     | 806317   | (M29067) unknown protein [Saccharomyces cerevisiae]                | 0.001   |
| 80     | Z49771                                | A.cepta mitochondrial gene for NADH dehydrogenase subunit 3 and ribosomal protein S12 | 4.5     | <NONE>   | <NONE>   | <NONE>  |
| 81     | M63494                                | Mouse IgG receptor (beta-Fc-gamma-RII) gene, exons 6 and 7, clones lambda-Fc(3.2,93). | 4.3     | <NONE>   | <NONE>   | <NONE>  |
| 82     | Z14035                                | S.pombe carl gene   | 2.0     | 3790665  | (AF099000) No definition line found [Caenorhabditis elegans]       | 1.2     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 83     | U17129                                | Rhodococcus erythropolis ThcA (thcA) gene, complete cds; and unknown genes        | 2.0     | 2828280  | (AL021687) putative protein [Arabidopsis thaliana]<br>>gi 2832633 gnl PID e1249651 (AL021711) putative protein [Arabidopsis thaliana] | 2e-26   |
| 84     | AE001386                              | Plasmodium falciparum chromosome 2, section 23 of 73 of the complete sequence     | 2.0     | 4176500  | (AL031177) dJ889M15.3 (novel protein)   | 9e-59   |
| 85     | U79292                                | Human clone 23734 mRNA sequence   | 1.9     | <NONE>   | <NONE>  | <NONE>  |
| 86     | V00159                                | Chloroplast Euglena gracilis gene coding for the 5S and 16S rRNA.                 | 1.9     | <NONE>   | <NONE>  | <NONE>  |
| 87     | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                           | 1.9     | <NONE>   | <NONE>  | <NONE>  |
| 88     | X93206                                | H.salinarium TATA box-binding protein genes and ORFs                              | 1.9     | <NONE>   | <NONE>  | <NONE>  |
| 89     | U60979                                | Caenorhabditis elegans programmed cell death specifier (ces-2) gene, complete cds | 1.9     | <NONE>   | <NONE>  | <NONE>  |
| 90     | X56272                                | C. tentans ORF's (A-E) for hemoglobin   | 1.9     | <NONE>   | <NONE>  | <NONE>  |
| 91     | L22383                                | Homo sapiens DNA sequence, repeat region.   | 1.9     | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 92     | U82814                                | Hirudo medicinalis neuron-specific protein mRNA, complete cds                                      | 1.9     | 3822533  | (AF094531) immunoglobulin heavy chain precursor   | 2.0     |
| 93     | U18504                                | Haplomitrium hookeri 18S rRNA gene, partial sequence.  | 1.9     | 1083969  | hypothetical protein 6 - fowlpox virus virus]   | 2.0     |
| 94     | X53676                                | Pseudomonas stutzeri nosDFY genes involved in copper processing                                    | 1.9     | 2980781  | (AL022198) putative protein   | 0.70    |
| 95     | U60086                                | Dictyostelium discoideum multidrug resistance transporter/Ser protease (tagC) mRNA, complete cds.  | 1.9     | 3879530  | (Z49130) cDNA EST yk486b9.3 comes from this gene; cDNA EST yk486b9.5 comes from this gene | 6e-05   |
| 96     | U33447                                | Human putative G-protein-coupled receptor (GPR17) gene; complete cds                               | 1.9     | 3880034  | (Z75550) similar to cell division control protein [Caenorhabditis elegans]                | 7e-14   |
| 97     | M81327                                | Sus scrofa lactoferrin mRNA, complete cds. > :: gb I28421 I28421 Sequence 5 from patent US 5571691 | 1.8     | <NONE>   | <NONE>  | <NONE>  |
| 98     | Y07622                                | S.iniae lctP & lctO genes and ORF1   | 1.8     | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 99     | M60474                                | Mouse myristoylated alanine-rich C-kinase substrate (MARCKS) mRNA, complete cds. | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 100    | Y13901                                | Homo sapiens FGFR-4 gene   | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 101    | U44400                                | Human Down Syndrome region of chromosome 21, clone A31D6-1D6.                    | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 102    | U92808                                | Ruminococcus albus beta-glucosidase (gluA) mRNA, complete cds                    | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 103    | L25051                                | Candida albicans argininosuccinate lyase (ARG4) gene, complete cds.              | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 104    | AE000546                              | Helicobacter pylori 26695 section 24 of 134 of the complete genome               | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 105    | J00978                                | Xenopus laevis major beta-globin gene, complete cds.                             | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 106    | U41716                                | Human immunodeficiency virus type 1 isolate JW95-5, vpr gene, complete cds.      | 1.8     | <NONE>   | <NONE>      | <NONE>  |
| 107    | X66286                                | G.gallus mRNA for tensin   | 1.8     | <NONE>   | <NONE>      | <NONE>  |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 108       | U76636                                | Xenopus calbindin D28k mRNA, complete cds                | 1.8     | <NONE>  | <NONE>   | <NONE>  |
| 109       | J00664                                | rabbit embryonic beta-4-globin gene.                     | 1.8     | <NONE>  | <NONE>   | <NONE>  |
| 110       | M21535                                | Human erg protein (ets-related gene) mRNA, complete cds. | 1.8     | 2983160   | (AE000693) hypothetical protein [Aquifex aeolicus]   | 7.7     |
| 111       | M80829                                | Rat troponin T cardiac isoform gene, complete cds        | 1.8     | 999450  | (Z46595) incomplete interleukin-11 receptor isoform [Homo sapiens]   | 7.3     |
| 112       | D37887                                | Cyprinus carpio c-myc gene for c-Myc, complete cds       | 1.8     | 3023408   | BRANCHED-CHAIN AMINO ACID TRANSPORT SYSTEM CARRIER PROTEIN (BRANCHED-CHAIN AMINO ACID UPTAKE CARRIER)<br>>gi 1075007 pir  D64056 membrane-associated component, branched amino acid transport system (brnQ) homolog - Haemophilus influenzae (strain Rd KW20) system II carrier protein (brnQ) [Haemophilus influenzae Rd] | 7.2     |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|---|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 113       | AF019765                              | Homo sapiens G protein-coupled receptor kinase 1 and G protein-coupled receptor kinase 1b (GRK1) gene, alternatively spliced, alternative exon 6, exon 7, and partial cds | 1.8     | 498643  | (U10270) G-box binding factor 1 [Zea mays]  | 7.2     |
| 114       | AF025967                              | Helicobacter pylori J166 virulence regulon transcriptional activator homolog gene, partial cds, strain-specific genomic sequence B2                                       | 1.8     | 3850108   | (AL033388) putative calcium-transporting atpase [Schizosaccharomyces pombe]   | 5.7     |
| 115       | U13183                                | Xenopus laevis (Xwnt-4) mRNA, complete cds.   | 1.8     | 2494853   | PROBABLE HYDROXYACYLGLUTATHIONE HYDROLASE (GLYOXALASE II) (GLX II) protein [Escherichia coli] >gi 1786406 (AE000130) probable hydroxyacylglutathione hydrolase [Escherichia coli] | 5.5     |
| 116       | S68944                                | Na <sup>+</sup> /Cl <sup>-</sup> -dependent neurotransmitter transporter  | 1.8     | 2276316   | (Z96810) GLYT-1 LIKE [Homo sapiens]   | 5.5     |
| 117       | M92905                                | Rat calcium channel alpha-1 subunit (rbB-I) mRNA, complete cds.   | 1.8     | 3165522   | (AF067607) Similar to cuticular collagen; C18H7.3   | 5.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 118    | X12429                                | Xenopus laevis U1 70K gene exon 10                               | 1.8     | 2735957  | (AF015685) reverse transcriptase domain protein   | 3.3     |
| 119    | D83333                                | Mouse hepatitis virus genomic RNA for spike protein, partial cds | 1.8     | 3876559  | (Z81072) Similarity to Human cyclin A/CDK2-associated protein P19 (RNA polymerase elongation factor) (SW:SKP1_HUMAN); cDNA EST EMBL:T00114 comes from this gene; cDNA EST yk390f11.5 comes from this gene; cDNA EST yk402e11.5 co... >gi 3877216 gnl PID e 1346850 protein P19 (RNA polymerase elongation factor) gene; cDNA EST yk390f11.5 comes from this gene; cDNA EST yk402e11.5 co... | 3.3     |
| 120    | AF016972                              | Cervus elaphus REDDEER mitochondrial D-loop, complete sequence   | 1.8     | 3878057  | (Z99942) similar to von Willebrand factor type A domain; cDNA EST yk412d4.5 comes from this gene; cDNA EST yk412d4.3 comes from this gene   | 3.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 121    | AB010741                              | Oncorhynchus mykiss mRNA for rtSox24, complete cds  | 1.8     | 1730805  | HYPOTHETICAL 21.0 KD PROTEIN IN RPS3-PSD1 INTERGENIC REGION<br>>gi 2132762 pir  S631 29 probable membrane protein YNL174w - yeast (Saccharomyces cerevisiae)<br>>gi 1302152 gnl PID e 239548 (Z71451) ORF YNL174w [Saccharomyces cerevisiae] | 2.5     |
| 122    | U32844                                | Haemophilus influenzae Rd section 159 of 163 of the complete genome   | 1.8     | 728910   | A-TYPE INCLUSION PROTEIN (ATI) camelpox virus<br>>gi 62381 (X69774) 84kDa A-type inclusion protein [unidentified]  | 1.9     |
| 123    | U18321                                | Human ionizing radiation resistance conferring protein mRNA, complete cds.  | 1.8     | 2133273  | ribosomal protein YS7 homolog Emericella nidulans  | 1.4     |
| 124    | M28668                                | Human cystic fibrosis mRNA, encoding a presumed transmembrane conductance regulator (CFTR). > :: gb I11500 I11500 Sequence 1 from Patent US 5407796 | 1.8     | 90492  | filaggrin precursor - mouse (fragment)   | 0.87    |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 125                                   | AF064553  | Mus musculus<br>NSD1 protein<br>mRNA,<br>complete cds            | 1.8     | 2501207   | PROBABLE<br>PROTEIN<br>DISULFIDE<br>ISOMERASE P5<br>PRECURSOR<br>>gi 1065461 (U40411)<br>Similar to protein<br>disulfide-isomerase.<br>[Caenorhabditis<br>elegans]       | 0.87    |
| 126                                   | AB002314  | Human mRNA<br>for KIAA0316<br>gene, complete<br>cds              | 1.8     | 115131  | REGULATORY<br>PROTEIN BRLA<br>(BRISTLE A<br>PROTEIN)<br>>gi 83718 pir  A28913<br>regulatory protein brlA<br>- Emericella nidulans<br>>gi 168029 (M20631)<br>brlA protein | 0.84    |
| 127                                   | L42096    | Homo sapiens<br>(subclone 10_d2<br>from P1 H21)<br>DNA sequence. | 1.8     | 2135624   | metalloproteinase 1<br>(EC 3.4.24.-) - human   | 0.65    |
| 128                                   | M37278    | R.norvegicus<br>renin gene,<br>exons 1-9.                        | 1.8     | 4050087   | (AF109907) S164<br>[Homo sapiens]  | 0.58    |
| 129                                   | X82879    | Artificial<br>sequences DNA<br>for ART 2<br>consensus            | 1.8     | 310929  | (L13442) cysteine-rich<br>extensin-like protein-4<br>[Nicotiana tabacum]   | 0.52    |
| 130                                   | D89729    | Homo sapiens<br>mRNA for<br>CRM1 protein,<br>complete cds        | 1.8     | 3559944   | (AJ010792) Muc5AC<br>protein [Mus<br>musculus]   | 0.38    |
| 131                                   | U78076    | Mus musculus<br>sepiapterin<br>reductase gene,<br>exons 1 and 2  | 1.8     | 2984225   | (AE000766) enolase-<br>phosphatase E-1<br>[Aquifex aeolicus]   | 0.095   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 132    | X52133                                | Paramecium 168G gene for 168G surface protein  | 1.8     | 115316   | COLLAGEN ALPHA I(VIII) CHAIN PRECURSOR (ENDOTHELIAL COLLAGEN)<br>>gi 105686 pir  S1543 5 collagen alpha 1(VIII) chain precursor -   | 0.073   |
| 133    | M77830                                | Human desmoplakin I mRNA, complete cds.  | 1.8     | 1397246  | (U61944) coded for by C. elegans cDNA yk112f3.5; coded for by C. elegans cDNA cm21d2; coded for by C. elegans cDNA CEESR07F; coded for by C. elegans cDNA yk112f3.3; coded for by C. elegans cDNA CEESR29F [Caenorhabditis elegans] | 1e-04   |
| 134    | AJ224150                              | Plasmodium berghei EF-lalpha A-gene  | 1.8     | 1353761  | (U43192) myosin II heavy chain [Naegleria fowleri]  | 2e-05   |
| 135    | AJ005518                              | Mus musculus somatostatin receptor 2 gene, exon1 and 5' flanking region              | 1.8     | 1326350  | (U58748) similar to potential transmembrane domains in S. cerevisiae nuclear division RFT1 protein (SP:P38206)  | 2e-08   |
| 136    | AF002217                              | Ralstonia eutropha megaplasmid pHG1 nitric oxide reductase (norB) gene, complete cds | 1.8     | 3393018  | (AL031174) hypothetical protein   | 2e-08   |
| 137    | AF039035                              | Caenorhabditis elegans cosmid C53A3  | 1.8     | 3850109  | (AL033388) 3-oxoacyl-[acyl-carrier-protein]-synthase  | 3e-11   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 138    | M81769                                | S.domesticus immunoglobulin rearranged gamma chain mRNA, VJC region, complete cds.                        | 1.8     | 3080527  | (AL022600) putative mannose-1 phosphate gaunyl transferase [Schizosaccharomyces pombe]  | 3e-14   |
| 139    | Y11106                                | P.pastoris PYC1 gene  | 1.8     | 1175412  | HYPOTHETICAL 24.2 KD PROTEIN C13A11.03 IN CHROMOSOME I >gi 984224 (Z54096) unknown  | 1e-15   |
| 140    | U87803                                | Human putative Ca2+/calmodulin-dependent protein kinase kinase gene, 3' flanking region, partial sequence | 1.8     | 2828280  | (AL021687) putative protein [Arabidopsis thaliana] >gi 2832633 gnl PID e 1249651 (AL021711) putative protein [Arabidopsis thaliana]   | 3e-17   |
| 141    | AE001430                              | Plasmodium falciparum chromosome 2, section 67 of 73 of the complete sequence                             | 1.8     | 1931647  | (U95973) endomembrane protein EMP70 precursor isolog  | 2e-20   |
| 142    | L19708                                | Rat N-methyl-D-aspartate receptor (NMDAR1) gene, first exon.  | 1.8     | 1731181  | HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II >gi 3874230 gnl PID e 1351618 protein (Swiss Prot accession number P38376); cDNA EST yk220e10.5 comes from this gene [Caenorhabditis elegans] | 3e-21   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 143       | Y10728                                | P.schwarzi<br>mitochondrial<br>cytb gene, partial  | 1.8     | 3878644   | (Z81103) predicted<br>using Genefinder;<br>cDNA EST<br>yk303g11.5 comes<br>from this gene; cDNA<br>EST yk303g11.3<br>comes from this gene<br>[Caenorhabditis<br>elegans] | 1e-28   |
| 144       | AB006631                              | Homo sapiens<br>mRNA for<br>KIAA0293 gene,<br>partial cds  | 1.8     | 4176500   | (AL031177)<br>dI889M15.3 (novel<br>protein)  | 7e-45   |
| 145       | AF106967                              | Mus musculus I3<br>protein mRNA,<br>complete cds   | 1.7     | <NONE>  | <NONE>   | <NONE>  |
| 146       | AE001073                              | Archaeoglobus<br>fulgidus section<br>34 of 172 of the<br>complete<br>genome  | 1.7     | <NONE>  | <NONE>   | <NONE>  |
| 147       | U12977                                | Pseudomonas<br>lemoignei<br>poly(3-<br>hydroxybutyrate<br>) depolymerase<br>A precursor<br>(phaZ5) gene,<br>complete cds,<br>and glycerol-3-<br>phosphate-<br>dehydrogenase<br>homolog,<br>complete cds. | 1.7     | <NONE>  | <NONE>   | <NONE>  |
| 148       | M27038                                | Mus musculus<br>(SK/CamRk)<br>germline IgK<br>chain gene, J1-5<br>region.  | 1.7     | <NONE>  | <NONE>   | <NONE>  |
| 149       | X74142                                | H.sapiens HBF-<br>1 mRNA for<br>transcripton<br>factor   | 1.7     | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|---|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 150    | U40830                                | Streptococcus thermophilus DeoD gene, partial cds and EpsA, EpsB, EpsC, EpsD, EpsE, EpsF, EpsG, EpsH, EpsI, EpsJ, EpsK, EpsL, EpsM, Orf14.9 protein genes, complete cds | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 151    | L29172                                | Rabbit Ig germline gamma H-chain (allotype d12,e15) C-region gene, 3' end.  | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 152    | M19045                                | Human lysozyme mRNA, complete cds.  | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 153    | AE001159                              | Borrelia burgdorferi (section 45 of 70) of the complete genome  | 1.7     | <NONE>   | <NONE>      | <NONE>  |
| 154    | L17027                                | Plasmid pFdA (from Fremyella diplosiphon) DNA sequence, including unidentified cds and stem loop.   | 1.7     | <NONE>   | <NONE>      | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|---------------------------------------|-----------|---|---------|---|-------------|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 155                                   | U12232    | Arabidopsis<br>thaliana<br>Columbia GTP<br>binding protein<br>beta subunit<br>(AGB1) mRNA,<br>complete cds. | 1.7     | <NONE>  | <NONE>      | <NONE>  |
| 156                                   | D42056    | Arabidopsis<br>thaliana ATPK6<br>mRNA for<br>ribosomal-<br>protein S6<br>kinase homolog,<br>complete cds    | 1.7     | <NONE>  | <NONE>      | <NONE>  |
| 157                                   | X98117    | Rhizobium<br>leguminosarum<br>prsD, prsE,<br>ORF3 genes   | 1.7     | <NONE>  | <NONE>      | <NONE>  |
| 158                                   | AF039084  | Spinacia<br>oleracea heat<br>shock 70 protein<br>protein,<br>complete cds                                   | 1.7     | <NONE>  | <NONE>      | <NONE>  |
| 159                                   | Z12651    | R.norvegicus<br>gene for catechol<br>methyltransferas<br>e  | 1.7     | <NONE>  | <NONE>      | <NONE>  |
| 160                                   | AF002968  | Fringilla coelebs<br>mitochondrial<br>control region,<br>partial sequence                                   | 1.7     | <NONE>  | <NONE>      | <NONE>  |
| 161                                   | AE001160  | Borrelia<br>burgdorferi<br>(section 46 of<br>70) of the<br>complete<br>genome                               | 1.7     | <NONE>  | <NONE>      | <NONE>  |
| 162                                   | U67553    | Methanococcus<br>jannaschii<br>section 95 of 150<br>of the complete<br>genome                               | 1.7     | <NONE>  | <NONE>      | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 163                                   | M86247    | S.ruminantium<br>plasmid pS23<br>DNA.                                   | 1.7     | <NONE>  | <NONE>   | <NONE>  |
| 164                                   | S74436    | oIL-<br>8=interleukin-8<br>[sheep, spleen<br>cells, mRNA,<br>1435 nt]   | 1.7     | <NONE>  | <NONE>   | <NONE>  |
| 165                                   | D12719    | Candida maltosa<br>ALK7<br>(CYP52A10)<br>and ALK8<br>complete cds       | 1.7     | <NONE>  | <NONE>   | <NONE>  |
| 166                                   | U02625    | Geotrichum<br>candidum NRRL<br>Y-553 lipase<br>gene, partial cds.       | 1.7     | 321245  | 230k bullous<br>pemphigoid antigen<br>BPM1 - mouse   | 9.3     |
| 167                                   | Z58881    | H.sapiens CpG<br>DNA, clone<br>114a4, reverse<br>read<br>cpg114a4.rtl . | 1.7     | 1854675   | (U66298) bone<br>morphogenetic<br>protein-6 [Rattus<br>norvegicus]   | 9.1     |
| 168                                   | U43674    | Agrobacterium<br>tumefaciens<br>conjugal transfer<br>region 1 genes     | 1.7     | 1352066   | LARGE PROLINE-<br>RICH PROTEIN<br>BAT2 MHC class III<br>histocompatibility<br>antigen HLA-B-<br>associated transcript 2<br>- human >gi 179339<br>(M33509) HLA-B-<br>associated transcript 2<br>(BAT2) [Homo<br>sapiens] >gi 179345<br>(M33518) HLA-B-<br>associated transcript 2<br>(BAT2) [Homo<br>sapiens] | 9.1     |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 169       | AL023827                              | Caenorhabditis elegans cosmid Y12A6A, complete sequence [Caenorhabditis elegans] | 1.7     | 731440  | PROTOPORPHYRIN OGEN OXIDASE (PPO) yeast (Saccharomyces cerevisiae) >gi 603606 (U18778) Hem14p: protoporphyrinogen oxidase [Saccharomyces cerevisiae] >gi 1403536 gnl PID e 249333 (Z71381) protoporphyrinogen oxidase [Saccharomyces cerevisiae] | 8.9     |
| 170       | X69662                                | X.laevis mRNA for glutathione synthetase, large subunit                          | 1.7     | 4038057   | (AC005897) hypothetical protein [Arabidopsis thaliana]   | 8.8     |
| 171       | Z35824                                | S.cerevisiae chromosome II reading frame ORF YBL063w                             | 1.7     | 3021450   | (Y15515) prdl-a [Hydra vulgaris]   | 7.0     |
| 172       | M65139                                | Cowpea chlorotic mottle virus (CCMV) 1a protein gene, complete cds.              | 1.7     | 2506307   | COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR 1(XII) chain - chicken >gi 222811 gnl PID d1 001160 gallus] >gi 2326442 gnl PID e 39435 (X61024) collagen type XII alpha 1 chain [Gallus gallus]   | 7.0     |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|--|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 173       | X15065                                | Drosophila distal<br>BX-C region<br>(bithorax<br>complex) pH189<br>5' region;    | 1.7     | 1723625   | HYPOTHETICAL<br>10.0 KD PROTEIN IN<br>ALPA-GABD<br>INTERGENIC<br>REGION (F87)<br>>gi 1033124 (U36840)<br>ORF_f87 [Escherichia<br>coli] >gi 1788982<br>(AE000348) orf,<br>hypothetical protein   | 6.9     |
| 174       | Z46255                                | S.cerevisiae<br>chromosome VI<br>lambda clone.                                   | 1.7     | 3875228   | (Z46792) similar to<br>lethal(1) discs large-1<br>tumor suppressor<br>protein-like repeats;<br>cDNA EST<br>EMBL:D33495 comes<br>from this gene; cDNA<br>EST EMBL:D35117<br>comes from this gene;<br>cDNA EST<br>EMBL:D36356 comes<br>from this gene; cDNA<br>EST EMB...<br>>gi 3879984 gnl PID e<br>1351767 suppressor<br>protein-like repeats;<br>cDNA EST<br>EMBL:D33495 comes<br>from this gene; cDNA<br>EST EMBL:D35117<br>comes from this gene;<br>cDNA EST<br>EMBL:D36356 comes<br>from this gene; cDNA<br>EST EMB... | 6.7     |
| 175       | U01066                                | Human CD4<br>promoter, partial<br>sequence.                                      | 1.7     | 125448  | THYMIDINE<br>KINASE saimiriine<br>herpesvirus 1 (strain<br>11[Onc]) >gi 60341   | 6.7     |
| 176       | U34743                                | Phalaenopsis sp.<br>'hybrid SM9108'<br>homeobox<br>protein mRNA,<br>complete cds | 1.7     | 1022918   | (U38184) ATPase<br>subunit 6<br>[Trypanosoma cruzi]   | 6.7     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 177    | U14662                                | Baboon herpesvirus HVP2 gB glycoprotein (UL27) gene, complete cds.                        | 1.7     | 3218378  | (AL023862) hypothetical protein SC3F9.07 [Streptomyces coelicolor]                                    | 6.7     |
| 178    | AB017006                              | Homo sapiens PMS2L15 mRNA, partial cds  | 1.7     | 1465855  | (U64859) glutamine-rich protein [Caenorhabditis elegans]  | 6.7     |
| 179    | U92651                                | Brassica oleracea var. botrytis tonoplast intrinsic protein bobTIP26-1 mRNA, complete cds | 1.7     | 3023675  | DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC) dynein heavy chain [Schizosaccharomyces pombe]                   | 6.6     |
| 180    | AF000634                              | Lytechinus variegatus notch homolog mRNA, complete cds                                    | 1.7     | 148574   | (M58520) endo-1,4-beta-glucanase [Fibrobacter succinogenes]   | 6.6     |
| 181    | M92354                                | Arabidopsis thaliana anthranilate synthase alpha subunit gene, complete cds.              | 1.7     | 738308   | blue light photoreceptor [Arabidopsis thaliana]   | 6.5     |
| 182    | AJ234856                              | Hordeum vulgare genomic DNA fragment; clone MWG2234.rev                                   | 1.7     | 3142302  | (AC002411) Strong similarity to myosin heavy chain gb Z34293 from A. thaliana. [Arabidopsis thaliana] | 6.5     |
| 183    | U76827                                | Stercorarius parasiticus bird J33 cytochrome b protein, partial cds                       | 1.7     | 3413810  | (Y17034) Bassoon [Mus musculus]   | 5.4     |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 184                                   | U05211    | Saccharomyces cerevisiae Ttp1p (TTP1) gene, complete cds.                             | 1.7     | 403173  | (L24492) lipoprotein [Rhodococcus erythropolis]  | 4.9     |
| 185                                   | AF076974  | Homo sapiens TRRAP protein (TRRAP) mRNA, complete cds                                 | 1.7     | 1170140   | PUTATIVE ENDOGLUCANASE TYPE K PRECURSOR (ENDO-1,4-BETA-GLUCANASE) (CELLULASE)  | 4.1     |
| 186                                   | AE000753  | Aquifex aeolicus section 85 of 109 of the complete genome                             | 1.7     | 1169357   | DNA ADENINE METHYLASE site-specific DNA-methyltransferase (adenine-specific) dam methylase gene product [Vibrio cholerae]  | 4.0     |
| 187                                   | AF005638  | Tupaia glis apolipoprotein AI prepropeptide mRNA, complete cds                        | 1.7     | 3355682   | (AL031124) putative secreted lyase   | 4.0     |
| 188                                   | M23090    | Human germline IgK chain gene V3-region, clone Humkv328h5                             | 1.7     | 2257483   | (AB004534) pi003 [Schizosaccharomyces pombe]   | 4.0     |
| 189                                   | M24001    | Mink enteritis virus antigenic type 2 capsid protein genes VP1 and VP2, complete cds. | 1.7     | 2143504   | myotonic dystrophy kinase - mouse (fragment) kinase, DM-kinase {C-terminal, alternatively spliced, clone delta II,III,IV,V} [mice, brain, Peptide Partial, 474 aa] [Mus sp.] | 3.9     |
| 190                                   | X59964    | H.sapiens CST4 gene for Cystatin D  | 1.7     | 1766075   | (U37273) winged helix protein CWH-2 [Gallus gallus]  | 3.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 191    | X95276                                | <i>P. falciparum</i> complete gene map of plastid-like DNA (IR-B)   | 1.7     | 3219951  | HYPOTHETICAL 11.7 KD PROTEIN C6B12.13 IN CHROMOSOME I >gi 2330843 gnl PID e334047 pombe] | 3.0     |
| 192    | D84487                                | Rat PMSG-induced ovarian mRNA, 3' sequence, N10   | 1.7     | 173164   | (J02719) valyl-tRNA synthetase [Saccharomyces cerevisiae]                                | 2.3     |
| 193    | L14851                                | <i>Rattus norvegicus</i> neurexin III-alpha gene, complete cds.   | 1.7     | 3323586  | (AF060869) single-strand binding protein [Salmonella typhimurium]                        | 2.3     |
| 194    | M97002                                | <i>Xenopus laevis</i> /gilli hybrid pseudo-IgH chain gene, V region, clone LG7G342A.  | 1.7     | 2118407  | MHC sex-limited protein - mouse (fragment) musculus]                                     | 2.3     |
| 195    | L07025                                | <i>Bacillus thuringiensis</i> delta-endotoxin (CryA(a)) gene, 5' end. > :: gb I34520 I34520 Sequence 1 from patent US 5596071 > :: gb I39790 I39790 Sequence 1 from patent US 5616495 > :: gb AR008487 AR008487 Sequence 1 from patent US 5753492 | 1.7     | 2496940  | HYPOTHETICAL 53.4 KD PROTEIN D1054.13 IN CHROMOSOME V >gi 3875316 gnl PID e1344967       | 1.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 196    | S73149                                | insulin-like growth factor II {intron 7} [human, Genomic, 1702 nt]                                | 1.7     | 3327038  | (AB014512) KIAA0612 protein [Homo sapiens]  | 1.8     |
| 197    | D86990                                | Human (lambda) DNA for immunoglobulin light chain   | 1.7     | 494367   | Fv Fragment (Murine Se155-4) Complex With The Trisaccharide: Alpha-D-Galactose(1-2)[alpha-D-Abequose(1-3)]alpha-D-Mannose (P1-Ome) (Part Of The Cell-Surface Carbohydrate Of Pathogenic Salmonella) | 1.8     |
| 198    | L17027                                | Plasmid pFdA (from Fremyella diplosiphon) DNA sequence, including unidentified cds and stem loop. | 1.7     | 1082702  | poliovirus receptor-related protein - human   | 1.4     |
| 199    | AL022273                              | Caenorhabditis elegans cosmid H22D14, complete sequence [Caenorhabditis elegans]                  | 1.7     | 3924605  | (AF069442) putative inhibitor of apoptosis [Arabidopsis thaliana]   | 1.4     |
| 200    | U89926                                | Drosophila melanogaster cut gene, partial sequence  | 1.7     | 2245100  | (Z97343) DNA-binding protein homolog  | 1.3     |
| 201    | Z25749                                | H.sapiens gene for ribosomal protein S7   | 1.7     | 2493459  | PROTEIN KINASE C SUBSTRATE, 60.1 KD PROTEIN, HEAVY CHAIN (PKCSH) (80K-H PROTEIN) >gij1215746  | 1.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 202    | U59841                                | Fundulus heteroclitus lactate dehydrogenase B                                     | 1.7     | 3005587  | (AF048977) Ser/Arg-related nuclear matrix protein [Homo sapiens]                      | 0.82    |
| 203    | X55763                                | Rabbit mRNA for smooth muscle calcium channel blocker (CaCB) receptor             | 1.7     | 3883128  | (AF082302) arabinogalactan-protein [Arabidopsis thaliana]                             | 0.82    |
| 204    | Z75528                                | Caenorhabditis elegans cosmid C18B12A, complete sequence [Caenorhabditis elegans] | 1.7     | 940397   | (D10123) core [Hepatitis C virus]   | 0.80    |
| 205    | U50912                                | Human XIST gene, poly purine-pyrimidine repeat region                             | 1.7     | 2338027  | (AF005370) large tegument protein [Alcelaphine herpesvirus 1]                         | 0.59    |
| 206    | X12817                                | Ovis aries beta-lactoglobulin gene  | 1.7     | 987050   | (X65335) lacZ gene product [unidentified cloning vector]                              | 0.45    |
| 207    | AF004419                              | Homo sapiens troponin T (TNNT2) gene, exon 13                                     | 1.7     | 2996364  | (AF053947) unknown [Yersinia pestis] >gi 3883090                                      | 0.22    |
| 208    | L43643                                | Gallus domesticus DNA microsatellite marker MCW119                                | 1.7     | 464896   | TRANSDUCIN-LIKE ENHANCER PROTEIN 1 enhancer-of-split homolog TLE-1 - human >gi 307510 | 0.20    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 209    | Z73278                                | S.cerevisiae chromosome XII reading frame ORF YLR106c   | 1.7     | 1351657  | HYPOTHETICAL 123.9 KD PROTEIN C30D11.04C IN CHROMOSOME I >gi 2130411 pir  S6256 2 hypothetical protein SPAC30D11.4c - fission yeast nuclear pore complex protein [Schizosaccharomyces pombe] | 0.20    |
| 210    | M22345                                | Mouse endogenous provirus gag, pol, and env region DNA.   | 1.7     | 2444455  | (AF020765) hypothetical protein [Myxococcus xanthus]   | 0.12    |
| 211    | AE000360                              | Escherichia coli K-12 MG1655 section 250 of 400 of the complete genome                              | 1.7     | 2736361  | (AF039038) No definition line found [Caenorhabditis elegans]   | 0.12    |
| 212    | AB020692                              | Homo sapiens mRNA for KIAA0885 protein, complete cds  | 1.7     | 2605924  | (AF029726) histidine kinase C [Dictyostelium discoideum]   | 0.094   |
| 213    | S69429                                | testis-determining gene/SRY homolog [Sminthopsis macroura=stripe d-faced dunnarts, Genomic, 855 nt] | 1.7     | 2499016  | TONB PROTEIN >gi 1666536 (U23764) TonB [Pseudomonas aeruginosa]  | 0.092   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 214    | S69429                                | testis-determining gene/SRY homolog [Sminthopsis macroura=striped-faced dunnarts, Genomic, 855 nt] | 1.7     | 2499016  | TONB PROTEIN >gi 1666536 (U23764) TonB [Pseudomonas aeruginosa]   | 0.088   |
| 215    | U67205                                | Mus musculus ACF7 neural isoform 3 (mACF7) mRNA, partial cds                                       | 1.7     | 2047349  | (AF000198) weak similarity to HSP90 [Caenorhabditis elegans]  | 0.052   |
| 216    | X98188                                | Artificial DNA sequence for mammalian lambda-neo minichromosome, 1400 bp                           | 1.7     | 2493779  | PUTATIVE CUTICLE COLLAGEN C09G5.6 collagen; cDNA EST yk244c3.5 comes from this gene; cDNA EST yk244c3.3 comes from this gene [Caenorhabditis elegans] | 0.042   |
| 217    | U70139                                | Mus musculus putative CCR4 protein mRNA, partial cds   | 1.7     | 2252630  | (U95973) hypothetical protein [Arabidopsis thaliana]  | 0.041   |
| 218    | L38808                                | Homo sapiens alpha-1 type V collagen (COL5A1) gene, 5' flank and exon 1.                           | 1.7     | 2895760  | (AF045246) universal minicircle sequence binding protein minicircle sequence binding protein [Crithidia fasciculata]                                  | 0.039   |
| 219    | Z72151                                | B.napus mRNA for AMP-binding protein   | 1.7     | 190475   | (K02576) salivary proline-rich protein 1 [Homo sapiens]   | 0.011   |
| 220    | X94152                                | R.norvegicus mRNA for cysteine sulfinic decarboxylase  | 1.7     | 2136212  | synapsin IIb - human >gi 1594277 (U40215) synapsin IIb [Homo sapiens]   | 0.008   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 221    | L20255                                | Mouse stathmin gene sequence.  | 1.7     | 2317934  | (U97553) unknown [murine herpesvirus 68]                                 | 0.006   |
| 222    | L13600                                | Rattus norvegicus glycine transporter mRNA, complete cds.              | 1.7     | 726403   | (U23175) similar to anion exchange protein [Caenorhabditis elegans]      | 0.003   |
| 223    | AJ224150                              | Plasmodium berghei EF-lalpha A-gene                                    | 1.7     | 2072290  | (U95094) XL-INCENP [Xenopus laevis]                                      | 0.001   |
| 224    | S80642                                | butyrophilin [mice, lactating mammary gland, mRNA Partial, 3193 nt]    | 1.7     | 2695746  | (AJ223010) Pmt2 [Schizosaccharomyces pombe]                              | 9e-04   |
| 225    | M22363                                | C.elegans unc-86 gene encoding two alternative proteins, complete cds. | 1.7     | 2224683  | (AB002369) KIAA0371 [Homo sapiens]                                       | 1e-04   |
| 226    | X92123                                | M.musculus cgt gene exon 1   | 1.7     | 3874232  | (Z49909) similar to Prokaryotic ribonuclease PH [Caenorhabditis elegans] | 3e-05   |
| 227    | AB016000                              | Ipomoea nil Pkn2 (knotted-like gene) mRNA, complete cds                | 1.7     | 2183083  | (AF000422) TTF-I interacting peptide 5 [Homo sapiens]                    | 1e-05   |



| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 228       | D14133                                | Bovine mRNA for synaptocanalin I   | 1.7     | 3925277   | (AL032643) similar to Uncharacterized protein family UPF0034, Double-stranded RNA binding motif; cDNA EST yk489b3.5 comes from this gene; cDNA EST yk439g7.5 comes from this gene [Caenorhabditis elegans] | 2e-06   |
| 229       | L01991                                | Mus musculus TAFG-1-like neuronal glycoprotein (PCS) mRNA, complete cds.       | 1.7     | 3006139   | (AL022299) hypothetical protein  | 4e-07   |
| 230       | X63016                                | Tomato yellow leaf curl virus Thailand isolate complete genome (TYLCV-THB-DNA) | 1.7     | 3643608   | (AC005395) hypothetical protein [Arabidopsis thaliana]   | 1e-07   |
| 231       | Z22802                                | H.sapiens microsatellite repeat. > :: gb G34562 G34562 human STS SHGC-51834    | 1.7     | 100210  | extensin precursor (clone Tom L 4) - tomato esculentum]  | 4e-09   |
| 232       | K02765                                | Human complement component C3 mRNA, alpha and beta subunits, complete cds.     | 1.7     | 2984320   | (AE000773) acetoin utilization protein [Aquifex aeolicus]  | 1e-09   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|--|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 233       | Z74818                                | S.cerevisiae<br>chromosome XV<br>reading frame<br>ORF YOL076w                | 1.7     | 3873700   | (Z73102) predicted<br>using Genefinder;<br>Similarity to Bacillus<br>subtilis DNAJ protein<br>gene; cDNA EST<br>EMBL:C12520 comes<br>from this gene; cDNA<br>EST EMBL:D71409<br>comes from this ge... | 7e-11   |
| 234       | D21871                                | Pig mRNA for<br>thimet<br>oligopeptidase                                     | 1.7     | 2632098   | (Y15513) Prodos<br>protein [Drosophila<br>melanogaster]   | 8e-13   |
| 235       | Y14344                                | Gallus gallus<br>gene encoding<br>neurofascin.<br>exons 9,10,11 &<br>12      | 1.7     | 3876421   | (Z81070) cDNA EST<br>EMBL:C12730 comes<br>from this gene; cDNA<br>EST yk200b6.5 comes<br>from this gene; cDNA<br>EST yk349a12.5<br>comes from this gene<br>[Caenorhabditis<br>elegans]                | 3e-14   |
| 236       | Z73608                                | S.cerevisiae<br>chromosome<br>XVI reading<br>frame ORF<br>YPL252c            | 1.7     | 1439663   | (U64605) C05D9.6<br>gene product<br>[Caenorhabditis<br>elegans]   | 6e-18   |
| 237       | AG000518                              | Homo sapiens<br>genomic DNA,<br>21q region,<br>clone: T171N23                | 1.7     | 1174468   | OLIGOSACCHARYL<br>TRANSFERASE<br>STT3 SUBUNIT<br>HOMOLOG<br>>gi 529357 (U13019)<br>No definition line<br>found [Caenorhabditis<br>elegans]  | 6e-18   |
| 238       | D17716                                | Human mRNA<br>for N-<br>acetylglucosami<br>nyltransferase V,<br>complete cds | 1.7     | 961446  | (D63877) KIAA0157<br>gene product is novel.   | 5e-19   |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 239                                   | AF102512  | Cheilodactylus vittatus country USA: Midway Island cytochrome c oxidase subunit I gene, mitochondrial gene encoding mitochondrial protein, partial cds | 1.7     | 1572756   | (U70848) C43G2.1 gene product [Caenorhabditis elegans] | 5e-40   |
| 240                                   | L30107    | Rattus norvegicus liver-specific transporter gene, promoter region.  | 1.7     | 4176443   | (AL022238) dJ1042K10.4 (novel protein)                 | 3e-49   |
| 241                                   | X91220    | H.sapiens mRNA for Na-Cl electroneutral thiazide-sensitive cotransporter   | 1.7     | 3478637   | (AC005546) R29425_1 [Homo sapiens]                     | 6e-54   |
| 242                                   | U97146    | Rattus norvegicus calcium-independent phospholipase A2 mRNA, complete cds  | 1.6     | <NONE>  | <NONE>   | <NONE>  |
| 243                                   | Z48508    | Pea seed-borne mosaic virus RNA for coat protein and polymerase (partial)  | 1.6     | <NONE>  | <NONE>   | <NONE>  |
| 244                                   | M18349    | Rat leukocyte common antigen (L-CA) gene, exons 1 through 5.   | 1.6     | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 245    | M13158                                | Yeast (S.pombe) cdc25+ gene (mitosis initiation), complete cds.               | 1.6     | <NONE>   | <NONE>   | <NONE>  |
| 246    | U39712                                | Mycoplasma genitalium section 34 of 51 of the complete genome                 | 1.6     | <NONE>   | <NONE>   | <NONE>  |
| 247    | M17922                                | Mouse Murine urokinase-type plasminogen activator protein gene, complete cds. | 1.6     | 3875750  | (Z81499) predicted using Genefinder; cDNA EST yk410e3.3 comes from this gene; cDNA EST yk410e3.5 comes from this gene [Caenorhabditis elegans] | 8.0     |
| 248    | M89986                                | Human polymorphic loci in Xq28.   | 1.6     | 3261710  | (Z84724) psd [Mycobacterium tuberculosis]  | 6.4     |
| 249    | M89986                                | Human polymorphic loci in Xq28.   | 1.6     | 2143805  | inositol-polyphosphate 4-phosphatase - rat   | 6.2     |
| 250    | U68725                                | Rattus norvegicus Deleted in colorectal Cancer                                | 1.6     | 1256804  | (U51449) RING3 protein [Xenopus laevis]  | 5.8     |
| 251    | X95199                                | P.platessa GSTA, GSTA1, GSTA2, and PPTN genes                                 | 1.6     | 3915113  | MALEYLACETATE REDUCTASE Pseudomonas cepacia >gi 643636 (U19883) maleylacetate reductase [Burkholderia cepacia]                                 | 4.9     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 252                                   | Y09103    | D.melanogaster RPA1 gene   | 1.6     | 3916021   | HYPOTHETICAL 91 KD PROTEIN IN COB INTRON<br>>gi 2654230 gnl PID e 1192341 (X02819)<br>unidentified reading frame<br>[Schizosaccharomyces pombe]  | 4.8     |
| 253                                   | Z14078    | T.aestivum mitochondrion fMet, 18S, 5S repeat unit DNA             | 1.6     | 2501668   | DYSTROPHIN-RELATED PROTEIN 2 sapiens]  | 3.6     |
| 254                                   | AB002314  | Human mRNA for KIAA0316 gene, complete cds                         | 1.6     | 130997  | REPETITIVE PROLINE-RICH CELL WALL PROTEIN 1 PRECURSOR<br>>gi 81809 pir  A29324<br>proline- rich protein precursor - soybean<br>>gi 170049 (J02746)<br>proline-rich protein [Glycine max] | 2.8     |
| 255                                   | M21488    | Human muscle creatine kinase gene (CKMM), exon 2.                  | 1.6     | 119399  | ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS:COAT PROTEIN GP62; COAT PROTEIN GP40]   | 2.2     |
| 256                                   | AE001164  | Borrelia burgdorferi (section 50 of 70) of the complete genome     | 1.6     | 4050089   | (AF109907)<br>hypothetical protein [Homo sapiens]  | 1.5     |
| 257                                   | X61757    | M.musculus rearranged T-cell receptor beta variable region (Vb17a) | 1.6     | 3377766   | (AF080090)<br>semaphorin IV isoform b [Mus musculus]   | 1.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 258    | M15346                                | T.cruzi tandemly repeated gene encoding an 85 kDa antigen with homology to heat shock proteins. | 1.6     | 2804437  | (AF043695) similar to zinc metalloprotease family of peptidases [Caenorhabditis elegans]                    | 0.41    |
| 259    | L39018                                | Rattus norvegicus sodium channel protein 6 (SCP6) mRNA, complete cds                            | 1.6     | 2920535  | (AF018081) type XVIII collagen [Homo sapiens]   | 0.037   |
| 260    | M29483                                | Human leukocyte adhesion protein p150,95 alpha subunit gene, exons 7 - 15.                      | 1.6     | 1840045  | (U49082) transporter protein [Homo sapiens]   | 2e-09   |
| 261    | L06844                                | Aspergillus niger beta-D-fructofuranosidase (suc1) gene, one exon.                              | 1.6     | 4206210  | (AF071527) putative calcium channel [Arabidopsis thaliana]  | 9e-10   |
| 262    | M10946                                | Chicken aldolase B gene, complete cds, clones lambda-C(11.1,4).                                 | 1.6     | 2746775  | (AF040640) similar to peptidase family C19 (ubiquitin carboxyl-terminal hydrolase) [Caenorhabditis elegans] | 1e-31   |
| 263    | X07881                                | Human gene PRB3L for proline-rich protein G1  | 1.5     | <NONE>   | <NONE>  | <NONE>  |
| 264    | U22260                                | Nicotiana tabacum UMP synthase (pyr5-6) mRNA, partial cds                                       | 1.5     | 3880923  | (Z99271) similar to Reverse transcriptase comes from this gene [Caenorhabditis elegans]                     | 0.50    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 265    | U76759                                | Mus musculus nuclear protein NIP45 mRNA, complete cds                          | 1.4     | 1330394  | (U58761) C01F1.6 gene product [Caenorhabditis elegans]  | 8.9     |
| 266    | AF076470                              | Rice tungro bacilliform virus Serdang strain, complete genome                  | 1.4     | 1703461  | POTASSIUM-TRANSPORTING ATPASE BETA CHAIN (PROTON PUMP) (GASTRIC H <sup>+</sup> /K <sup>+</sup> ATPASE BETA SUBUNIT) 3.6.1.36) beta chain - human >gi 184105 (M75110) H,K-ATPase beta subunit [Homo sapiens] | 8.9     |
| 267    | X64659                                | C.jacchus interferon gene for interferon gamma                                 | 1.4     | 1486485  | (U28832) US10 [Gallid herpesvirus 1] >gi 1486497  | 6.8     |
| 268    | U11825                                | Schistosoma japonicum structural muscle protein paramyosin mRNA, complete cds. | 0.88    | <NONE>   | <NONE>  | <NONE>  |
| 269    | D84278                                | Human DNA for CD38, exon 1   | 0.68    | 3766363  | (AL031907) hypothetical serine rich protein [Schizosaccharomyces pombe]   | 3.0     |
| 270    | M59755                                | Bovine lens aldose reductase pseudogene, 3' end.                               | 0.67    | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 271    | M81758                                | Homo sapiens skeletal muscle voltage-dependent sodium channel alpha subunit (SkM1) mRNA, complete cds.  | 0.65    | 2437819  | (Z86105) 1,4-beta-glucanase [Anaerocellum thermophilum]  | 3.6     |
| 272    | L01965                                | Human type IV sodium channel alpha polypeptide  | 0.64    | 2437819  | (Z86105) 1,4-beta-glucanase [Anaerocellum thermophilum]  | 3.5     |
| 273    | U90122                                | Danio rerio bone morphogenetic protein-4 (bmp4) mRNA, partial cds                                       | 0.63    | 2983532  | (AE000720) formate dehydrogenase alpha subunit [Aquifex aeolicus]  | 7.9     |
| 274    | L41624                                | Hylobates lar mucin (MUC1) gene, exons 1-6.   | 0.63    | 1517808  | (D79215) FGF-10 [Rattus norvegicus]  | 0.91    |
| 275    | AF030881                              | Fugu rubripes sushi retrotransposon gag polyprotein (gag) and pol polyprotein (pol) genes, complete cds | 0.63    | 1519696  | (U67956) coded for by C. elegans cDNA yk126f9.5; coded for by C. elegans cDNA yk159h6.3; coded for by C. elegans cDNA yk126f9.3; coded for by C. elegans cDNA yk159h6.5 [Caenorhabditis elegans] | 0.38    |
| 276    | U52909                                | Arabidopsis thaliana U1 snRNP 70K protein gene, complete cds  | 0.62    | <NONE>   | <NONE>   | <NONE>  |
| 277    | AF008192                              | Homo sapiens putative GR6 protein (GR6) mRNA, complete cds  | 0.62    | 3800934  | (AF100655) contains similarity to ser/thr protein kinases [Caenorhabditis elegans]   | 9.7     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 278    | U17081                                | Human fatty acid binding protein (FABP3) gene, complete cds                 | 0.62    | 3617848  | (AF049709) tyrosylprotein sulfotransferase-A; TPST-A  | 7.7     |
| 279    | AB018340                              | Homo sapiens mRNA for KIAA0797 protein, partial cds                         | 0.62    | 424044   | VP5 protein - porcine rotavirus >gi 61355   | 7.7     |
| 280    | Y00093                                | H.sapiens mRNA for leukocyte adhesion glycoprotein p150,95                  | 0.62    | 1054945  | (U38621) polyprotein [Tobacco vein mottling virus]  | 4.5     |
| 281    | M63138                                | Human cathepsin D (catD) gene, exons 7, 8, and 9.                           | 0.62    | 136810   | GLYCOPROTEIN M >gi 73791 pir  WMBE 51 UL10 protein - human herpesvirus 1 1-473) [Human herpesvirus 1] >gi 221732 gnl PID d1 002131                | 3.5     |
| 282    | X76056                                | N. sylvestris DNA for spacer region between 25S and 18S ribosomal RNA genes | 0.62    | 2661176  | (U76671) putative cds [Rhodobacter sphaeroides]   | 2.0     |
| 283    | X74501                                | B.taurus mRNA for ACTH receptor   | 0.62    | 4249552  | (AB001075) galectin-2 related protein   | 2.0     |
| 284    | M57634                                | Rat F1-ATPase beta subunit mRNA, 3' end.                                    | 0.62    | 2119692  | transforming growth factor-beta type III receptor - chicken >gi 511843 (L01121) transforming growth factor-beta type III receptor [Gallus gallus] | 1.5     |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 285       | Y15724                                | Homo sapiens<br>SERCA3 gene,<br>exons 1-7 (and<br>joined CDS)  | 0.62    | 2498164   | ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-DIOXYGENASE) beta-dioxygenase (EC 1.14.11.16) - bovine >gi 162694 taurus]                  | 0.52    |
| 286       | AL010142                              | Plasmodium falciparum<br>DNA ***<br>SEQUENCING<br>IN<br>PROGRESS ***<br>from contig 3-72, complete<br>sequence | 0.62    | 3183206   | HYPOTHETICAL<br>PROTEIN KIAA0009<br>sapiens]   | 4e-07   |
| 287       | AB008160                              | Mus musculus<br>Stat3 gene, 5'-<br>flanking region<br>and exon 1<br>partial sequence                           | 0.62    | 466097  | HYPOTHETICAL<br>63.5 KD PROTEIN<br>ZK353.1 IN<br>CHROMOSOME III<br>>gi 1078903 pir  S446<br>54 ZK353.1 protein -<br>Caenorhabditis<br>elegans >gi 289757<br>(L15313) putative<br>[Caenorhabditis<br>elegans] | 1e-35   |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 288                                   | AB018795  | Halomonas marina gene for alginate lyase, complete cds                                    | 0.62    | 3877493   | (Z48583) similar to ATPases associated with various cellular activities (AAA); cDNA EST EMBL: Z14623 comes from this gene; cDNA EST EMBL:D75090 comes from this gene; cDNA EST EMBL:D72255 comes from this gene; cDNA EST yk200e4... | 3e-46   |
| 289                                   | Z69906    | Human DNA sequence from cosmid E141E2, on chromosome 22, complete sequence [Homo sapiens] | 0.61    | <NONE>  | <NONE>   | <NONE>  |
| 290                                   | U18259    | Human clone CIITA-8 MHC class II transactivator CIITA mRNA, complete cds.                 | 0.61    | 1483567   | (X79983) viral proteinase [Pseudorabies virus]   | 9.8     |
| 291                                   | X98890    | S.tuberosum mRNA for inorganic phosphate transporter, StPT1                               | 0.61    | 475724  | (U08884) protein VIII precursor [Bovine adenovirus type 3]   | 7.6     |
| 292                                   | U70825    | Rattus norvegicus 5-oxo-L-prolinase mRNA, complete cds                                    | 0.61    | 733543  | (U23448) similar to genome polypeptide (SP:POLG_BVDVN, P19711); alternative splicing to C04A2.7a   | 4.4     |
| 293                                   | L81667    | Homo sapiens (subclone 2_a9 from P1 H49) DNA sequence                                     | 0.61    | 2565087   | (U80759) CAGH4 alternate open reading frame [Homo sapiens]   | 3.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 294    | AE000760                              | Aquifex aeolicus section 92 of 109 of the complete genome                                      | 0.61    | 2811092  | HOMEBOX PROTEIN HOX-A3 (HOX-1.5) homeobox-containing transcription factor [Mus musculus]                         | 2.6     |
| 295    | U58512                                | Mus musculus Rho-associated, coiled-coil forming protein kinase p160 ROCK-1 mRNA, complete cds | 0.61    | 295671   | (L11275) selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III | 1.5     |
| 296    | U27459                                | Human origin recognition complex protein 2 homolog hORC2L mRNA, complete cds                   | 0.61    | 200285   | (M97900) putative open reading frame [Mus musculus]  | 0.66    |
| 297    | L36680                                | Pisum sativum S-adenosylmethionine synthase mRNA, 3' end.                                      | 0.61    | 2285790  | (AB002086) p47 [Rattus norvegicus]   | 4e-12   |
| 298    | AE000673                              | Aquifex aeolicus section 5 of 109 of the complete genome                                       | 0.61    | 3395782  | (AF058446) histone macroH2A1.2 [Gallus gallus]   | 6e-27   |
| 299    | AF086310                              | Homo sapiens full length insert cDNA clone ZD51F08   | 0.61    | 3646450  | (AL031603) conserved hypothetical protein. [Schizosaccharomyces pombe]   | 8e-29   |
| 300    | AJ009675                              | Agrotis ipsilon mRNA for 3-hydroxy-3-methylglutaryl coenzyme A reductase                       | 0.61    | 4176370  | (AC005058) similar to calcium-independent phospholipase A2; similar to AC004392 (PID:g3367519) [Homo sapiens]    | 2e-73   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|---|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 301    | AC005577                              | Homo sapiens chromosome 19, cosmid F18382B, centromeric end, complete sequence [Homo sapiens]   | 0.60    | <NONE>   | <NONE>      | <NONE>  |
| 302    | U40454                                | Candida albicans topoisomerase type I (CATOP1) gene, complete cds   | 0.60    | <NONE>   | <NONE>      | <NONE>  |
| 303    | J01390                                | Emericella nidulans mtDNA between h2/h5 and bh2/b2 junctions. genes for ATPase subunit 6, cytochrome oxidase subunit 3, seven. unidentified proteins, twentyfour tRNA's and L-rRNA. | 0.60    | <NONE>   | <NONE>      | <NONE>  |
| 304    | L11172                                | Plasmodium falciparum RNA polymerase I gene, complete cds.  | 0.60    | <NONE>   | <NONE>      | <NONE>  |
| 305    | Z81079                                | Caenorhabditis elegans cosmid F39H11, complete sequence [Caenorhabditis elegans]  | 0.60    | <NONE>   | <NONE>      | <NONE>  |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|--|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 306       | Z49627                                | S.cerevisiae<br>chromosome X<br>reading frame<br>ORF YJR127c   | 0.60    | 118751  | MAJOR DNA-<br>BINDING PROTEIN<br>herpesvirus 1 (strain<br>11) >gi 60327<br>(X64346) major<br>ssDNA-binding<br>protein [Saimiriine<br>herpesvirus 2] | 9.6     |
| 307       | U94911                                | Rattus<br>norvegicus H-K-<br>ATPase alpha 2<br>gene,<br>alternatively<br>spliced products<br>and partial cds | 0.60    | 2213862   | (AF003086) PfSNF2L<br>[Plasmodium<br>falciparum]  | 7.4     |
| 308       | U67476                                | Methanococcus<br>jannaschii<br>section 18 of 150<br>of the complete<br>genome                                | 0.60    | 1749688   | (D89240) unnamed<br>protein product   | 5.7     |
| 309       | U67513                                | Methanococcus<br>jannaschii<br>section 55 of 150<br>of the complete<br>genome                                | 0.60    | 3327421   | (U97068) zonadhesin<br>[Mus musculus]   | 4.3     |
| 310       | U57817                                | Haemophilus<br>ducreyi<br>lipoprotein gene,<br>complete cds  | 0.60    | 4008577   | (AL034491)<br>conserved hypothetical<br>protein<br>[Schizosaccharomyces<br>pombe]   | 2.5     |
| 311       | X80700                                | H.sapiens G17<br>gene  | 0.60    | 422541  | probable protein-<br>tyrosine kinase (EC<br>2.7.1.112) RTK -<br>Pacific electric ray<br>>gi 290858  | 1.5     |
| 312       | L42167                                | Mus musculus<br>(clone R24) rds<br>gene, partial cds   | 0.60    | 4220848   | (AF033823) moira<br>[Drosophila<br>melanogaster]  | 0.51    |
| 313       | U54777                                | Human hMSH6<br>mRNA,<br>complete cds   | 0.60    | 2665637   | (AF031087) mismatch<br>repair protein MSH6<br>[Mus musculus]  | 5e-07   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 314    | D86985                                | Human mRNA for KIAA0232 gene, complete cds                               | 0.60    | 1938462  | (U97006) No definition line found [Caenorhabditis elegans]  | 2e-07   |
| 315    | D43964                                | Rat liver mRNA for Kan-1, complete cds                                   | 0.60    | 1280135  | (U55376) coded for by C. elegans cDNA cm21e6; coded for by C. elegans cDNA cm01e2; similar to melibiose carrier protein (thiomethylgalactoside permease II) | 5e-15   |
| 316    | U49058                                | Rattus norvegicus CTD-binding SR-like protein rA4 mRNA, partial cds      | 0.60    | 2145091  | (U37500) RNA polymerase II largest subunit [Mus musculus]   | 1e-19   |
| 317    | X84388                                | U.ruddi mitochondrial 12S ribosomal RNA                                  | 0.60    | 3874247  | (Z70205) predicted using Genefinder   | 2e-37   |
| 318    | AF125447                              | Caenorhabditis elegans cosmid Y14H12B                                    | 0.59    | <NONE>   | <NONE>  | <NONE>  |
| 319    | U20189                                | Hyoscyamus muticus clone cVS2 vetispiradiene synthase mRNA, partial cds. | 0.59    | <NONE>   | <NONE>  | <NONE>  |
| 320    | M63962                                | Human gastric H,K-ATPase catalytic subunit gene, complete cds.           | 0.59    | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 321    | AJ132366                              | Helicobacter pylori (strain P1) comB and pmi/algA (partial) genes, and partial ORF1 and ORF2         | 0.59    | <NONE>   | <NONE>   | <NONE>  |
| 322    | U17289                                | Mus musculus transcription factor AP-2 (AP-2) gene, alternative exon 1a, and isoform 2, partial cds. | 0.59    | 2459419  | (AC002332) hypothetical protein [Arabidopsis thaliana]                     | 9.4     |
| 323    | Z71466                                | S.cerevisiae chromosome XIV reading frame ORF YNL190w  | 0.59    | 3875542  | (Z67990) Similarity to Rat amiloride-sensitive sodium channel beta-subunit | 7.3     |
| 324    | Z66493                                | Beet soil-borne virus genes for 13K, 22K and 48K proteins  | 0.59    | 2119867  | cryV465 protein - Bacillus thuringiensis thuringiensis]                    | 7.2     |
| 325    | L41351                                | Homo sapiens prostatic mRNA, complete cds  | 0.59    | 729212   | CRYSTALLIN J1C crystallin [Tripedalia cystophora]                          | 4.2     |
| 326    | X79854                                | S.lincolnensis gene for 16S ribosomal RNA  | 0.59    | 3702828  | (AF056577) high mobility group protein 1.2                                 | 3.2     |
| 327    | AJ223356                              | Strongylocentrotus purpuratus mRNA for SuDp98 protein  | 0.59    | 2495704  | HYPOTHETICAL PROTEIN KIAA0129 product is novel. [Homo sapiens]             | 2.5     |
| 328    | X86019                                | H.sapiens mRNA for PRPL-2 protein  | 0.59    | 1743341  | (Y10027) transcription factor TEF-1 [Mus musculus]                         | 2.5     |
| 329    | U75528                                | Xiphias gladius creatine kinase gene, partial cds  | 0.59    | 1845995  | (U69477) envelope glycoprotein [Human immunodeficiency virus type 1]       | 2.4     |



| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|---|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 330       | AC005573                              | Homo sapiens<br>chromosome 5,<br>PAC clone<br>202e13                                | 0.59    | 2506366   | DNA POLYMERASE<br>EPSILON SUBUNIT<br>B DNA-directed DNA<br>polymerase (EC<br>2.7.7.7) II chain B -<br>yeast ( <i>Saccharomyces<br/>cerevisiae</i> ) >gi 786319<br>(U25842) DNA<br>Polymerase epsilon,<br>subunit B (Swiss Prot.<br>accession number<br>P24482)<br>[ <i>Saccharomyces<br/>cerevisiae</i> ] | 1.4     |
| 331       | L19180                                | Rat receptor-<br>linked protein<br>tyrosine<br>phosphatase                          | 0.59    | 1235974   | (X96713) collagen<br>[ <i>Globodera pallida</i> ]   | 1.1     |
| 332       | L32090                                | <i>Listeria<br/>monocytogenes</i><br>secA gene,<br>complete cds.                    | 0.59    | 2291129   | (AF016415) No<br>definition line found<br>[ <i>Caenorhabditis<br/>elegans</i> ]   | 0.83    |
| 333       | U24433                                | <i>Xenopus laevis</i><br>syndecan-2<br>mRNA,<br>complete cds.                       | 0.59    | 3355692   | (AL031124)<br>hypothetical protein<br>SC1C2.25c<br>[ <i>Streptomyces<br/>coelicolor</i> ]   | 0.64    |
| 334       | M23412                                | <i>Drosophila</i><br>muscarinic<br>acetylcholine<br>receptor mRNA,<br>complete cds. | 0.59    | 168237  | (M76546)<br>hydroxyproline-rich<br>protein [ <i>Helianthus<br/>annuus</i> ]   | 0.22    |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 335                                   | AF060729  | Synaphea media chloroplast atpB-rbcL intergenic spacer region, partial sequence  | 0.59    | 731596  | HYPOTHETICAL<br>67.5 KD PROTEIN IN PRPS4-STE20 INTERGENIC REGION<br>>gi 626567 pir  S4682<br>5 hypothetical protein YHL010c - yeast (Saccharomyces cerevisiae)<br>>gi 2289881 (U11582)<br>No definition line found [Saccharomyces cerevisiae]      | 0.16    |
| 336                                   | AF029734  | Xanthobacter autotrophicus transcriptional activator AldR (aldR) gene, partial cds; and NAD-dependent chloroacetaldehyde dehydrogenase (aldB) gene, complete cds | 0.59    | 2498801   | PERIAXIN<br>>gi 2143901 pir  I5815<br>7 periaxin - rat<br>>gi 505297 (Z29649)<br>periaxin [Rattus norvegicus]  | 0.13    |
| 337                                   | X95307    | C.reinhardtii LI818r-l gene  | 0.59    | 1723781   | HYPOTHETICAL<br>34.3 KD PROTEIN IN TAF145-YOR1 INTERGENIC REGION<br>>gi 2131717 pir  S646<br>12 hypothetical protein YGR277c - yeast (Saccharomyces cerevisiae)<br>>gi 1323505 gnl PID e<br>243248 (Z73062) ORF YGR277c [Saccharomyces cerevisiae] | 1e-04   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|---|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 338       | M24572                                | Dictyostelium discoideum tRNA-Glu-GAA gene, clone yGluGAA8.           | 0.59    | 1176186   | HYPOTHETICAL 43.3 KD GTP-BINDING PROTEIN IN DACB-RPMA INTERGENIC REGION >gi 606121 coli] | 3e-06   |
| 339       | U73733                                | Human hMSH6 gene, exon 2  | 0.59    | 2665637   | (AF031087) mismatch repair protein MSH6 [Mus musculus]                                   | 5e-07   |
| 340       | D90747                                | Escherichia coli genomic DNA. (25.2 - 25.6 min)                       | 0.59    | 134286  | DOLICHOL KINASE  | 6e-08   |
| 341       | J05211                                | Human desmoplakin mRNA, 3' end.                                       | 0.59    | 246796  | major centromere protein, CENP-B [human, Peptide, 594 aa]                                | 4e-08   |
| 342       | L24441                                | Loligo pealii kinesin light chain mRNA, complete cds.                 | 0.59    | 547800  | KINESIN LIGHT CHAIN (KLC) sea urchin (Strongylocentrotus purpuratus) >gi 161530          | 5e-14   |
| 343       | M25140                                | Human cardiac alpha-myosin heavy chain (MYH6) gene, exons 2, 3 and 4. | 0.58    | <NONE>  | <NONE>   | <NONE>  |
| 344       | L81932                                | Homo sapiens (subclone 9_h2 from P1 H21) DNA sequence                 | 0.58    | <NONE>  | <NONE>   | <NONE>  |
| 345       | AF087966                              | Homo sapiens full length insert cDNA clone YU51G04                    | 0.58    | <NONE>  | <NONE>   | <NONE>  |
| 346       | Z78574                                | H.sapiens flow-sorted chromosome 6 TaqI fragment, SC6pA10G11          | 0.58    | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 347    | AF068061                              | Blattella germanica allatostatin neuropeptide precursor, gene, complete cds  | 0.58    | <NONE>   | <NONE>   | <NONE>  |
| 348    | AF015592                              | Homo sapiens Cdc7 (CDC7) mRNA, complete cds  | 0.58    | <NONE>   | <NONE>   | <NONE>  |
| 349    | AF028006                              | Methanosarcina barkeri atp operon: ATP synthase beta subunit (atpD), ATP synthase epsilon subunit (atpC), ATP synthase gene 1 (atpI), ATP synthase a subunit subunit (...) | 0.58    | 3184291  | (AC004136) putative DNA polymerase III gamma subunit | 9.4     |
| 350    | AB017032                              | Mus musculus gene for pancreatic trypsin, complete cds   | 0.58    | 3170561  | (AF056704) synapsin IIIa [Rattus norvegicus]         | 9.2     |
| 351    | AF081585                              | Dictyostelium discoideum developmental protein DG1110 (DG1110) gene, partial cds   | 0.58    | 105417   | basic proline-rich peptide IB-8a human               | 9.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 352    | AF086322                              | Homo sapiens full length insert cDNA clone ZD53E01   | 0.58    | 93026  | hypothetical protein - African swine fever virus (strain Malawi Lil-20/1) >gi 450758 (X71982) myeloid differentiation antigen homologue [African swine fever virus] >gi 903686 (M95672) unknown protein | 7.1     |
| 353    | AF088025                              | Homo sapiens full length insert cDNA clone ZC19C04   | 0.58    | 2384644  | (U92805) thrombospondin-3 [Xenopus laevis]  | 7.0     |
| 354    | AB002339                              | Human mRNA for KIAA0341 gene, partial cds  | 0.58    | 2135587  | M130 antigen (cytosolic variant 2) - human  | 5.4     |
| 355    | U67548                                | Methanococcus jannaschii section 90 of 150 of the complete genome  | 0.58    | 2911094  | (AL021957) hypothetical protein Rv2174  | 4.2     |
| 356    | L07868                                | Homo sapiens receptor tyrosine kinase (ERBB4) gene, complete cds.  | 0.58    | 461922   | PYRUVATE DECARBOXYLASE (8-10 NM CYTOPLASMIC FILAMENT-ASSOCIATED PROTEIN) (P59NC) 4.1.1.1) - Neurospora crassa >gi 293948 (L09125) pyruvate decarboxylase [Neurospora crassa] >gi 1655909                | 4.2     |
| 357    | X03897                                | Bacillus subtilis sigma 43 operon with P23-dnaE-rpoD genes (dnaE for DNA primase, rpoD for RNA polymerase) | 0.58    | 1323704  | (U55387) similar to C. elegans F38E1.9 gene product encoded by GenBank Accession Number U41996 [Cricetulus griseus]   | 4.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 358    | D76419                                | Desulfovibrio vulgaris rbo gene for desulfoferrodoxin and rub gene for rubredoxin, complete cds | 0.58    | 3420047  | (AC004680) putative protein kinase [Arabidopsis thaliana]   | 2.4     |
| 359    | Z82174                                | Human DNA sequence from cosmid B20F6 on chromosome 22, complete sequence [Homo sapiens]         | 0.58    | 2145455  | (Y07866) catalase-peroxidase  | 2.4     |
| 360    | M33642                                | F.solani STI35 protein gene, complete cds.  | 0.58    | 2896706  | (AL021897) hypothetical protein Rv1069c   | 2.4     |
| 361    | U64873                                | Mus musculus transforming growth factor alpha (TGF alpha) gene, partial cds                     | 0.58    | 3874437  | (Z81038) predicted using Genefinder; cDNA EST yk488a2.5 comes from this gene [Caenorhabditis elegans] | 1.8     |
| 362    | AB002132                              | Macrophthalmus banzai mitochondrial DNA for 12S and 16S rRNA, partial and complete sequence     | 0.58    | 2960022  | (AJ224676) rho type GEF [Drosophila melanogaster]   | 1.8     |
| 363    | AF070070                              | Caenorhabditis elegans MutS homolog (msh-5) mRNA, partial cds                                   | 0.58    | 4098205  | (U75869) Omp22 [Helicobacter pylori]  | 1.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 364    | AF045240                              | Staphylococcus epidermidis plasmid pIP1629 mobilization protein (mobC1), (orf69-1), (mobA1),  | 0.58    | 4218117  | (AL035353) protein (fragment)   | 0.62    |
| 365    | X61637                                | H.sapiens Wilms tumor gene 1, exons 8 and 9   | 0.58    | 2331059  | (U88211) unknown [Gallus gallus]  | 0.62    |
| 366    | AF039312                              | Moraxella catarrhalis strain 4223 transferrin binding protein A (tbpA) and transferrin binding protein B (tbpB) genes, complete cds; and unknown gene | 0.58    | 120155   | FIBER PROTEIN >gi 74229 pir  ERADF M fiber protein - mouse adenovirus 1 >gi 209758 (M30594) fiber protein [Mastadenovirus mus1] | 0.27    |
| 367    | D87463                                | Human mRNA for KIAA0273 gene, complete cds  | 0.58    | 3861477  | (U94177) androgen receptor [Pan troglodytes]  | 0.12    |
| 368    | U40342                                | Mus musculus ninein mRNA, complete cds.   | 0.58    | 4115936  | (AF118223) No definition line found [Arabidopsis thaliana]  | 0.004   |
| 369    | S57235                                | CD68=110kda transmembrane glycoprotein [human, promonocyte cell line U937, mRNA, 1722 nt]   | 0.58    | 2072501  | (U96113) WWP1 [Homo sapiens]  | 1e-04   |
| 370    | U39391                                | Mus musculus serotonin1A receptor mRNA, complete cds.   | 0.58    | 1469876  | (D63481) The KIAA0147 gene product is related to adenylyl cyclase. [Homo sapiens]   | 1e-07   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 371    | D00056                                | Monkey B-lymphotropic papovavirus genes for VP-1, 2, 3 and large T antigen, complete and partial cds, strain LPV-76 > :: gb M14494 PPM VP1M Monkey B-lymphotropic papovavirus mutant (LPV-76) PstI B fragment encoding VP1, VP2, VP3 and T-antigen. | 0.58    | 2462069  | (AJ001774) vanadium chloroperoxidase   | 1e-08   |
| 372    | M77182                                | Amsacta entomopoxvirus spheroidin gene, complete cds, and four vaccinia related orfs. > :: gb I16670 I16670 Sequence 1 from patent US 5476781   | 0.58    | 1730722  | HYPOTHETICAL 43.8 KD PROTEIN IN NCE3-HHT2 INTERGENIC REGION >gi 2131871 pir S629 57 hypothetical protein YNL035c - yeast (Saccharomyces cerevisiae) >gi 1301880 gnl PID e 239670 (Z71311) ORF YNL035c [Saccharomyces cerevisiae] | 8e-14   |
| 373    | S72579                                | igloo-S=growth-associated protein GAP-43 homolog  | 0.58    | 2689720  | (AF037168) DnaJ homologue [Arabidopsis thaliana]   | 7e-14   |
| 374    | AF018165                              | Tetraodon fluviatilis amyloid precursor protein mRNA, complete cds  | 0.58    | 3219938  | HYPOTHETICAL 34.9 KD PROTEIN C57A10.11C IN CHROMOSOME I >gi 2058378 gnl PID e 314002 pombe]  | 5e-22   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 375    | U81803                                | Filobasidiella neoformans translation elongation factor EF1-alpha (CnTEF1) mRNA, complete cds        | 0.57    | <NONE>   | <NONE>      | <NONE>  |
| 376    | U09781                                | Candida albicans ATCC 18804, CBS 562 peptide transporter gene, complete cds.                         | 0.57    | <NONE>   | <NONE>      | <NONE>  |
| 377    | AC002143                              | Homo sapiens (subclone 4_b10 from BAC H102) DNA sequence   | 0.57    | <NONE>   | <NONE>      | <NONE>  |
| 378    | U23442                                | Tetrahymena thermophila RR internal deletion sequence.   | 0.57    | <NONE>   | <NONE>      | <NONE>  |
| 379    | U17289                                | Mus musculus transcription factor AP-2 (AP-2) gene, alternative exon 1a, and isoform 2, partial cds. | 0.57    | <NONE>   | <NONE>      | <NONE>  |
| 380    | X70844                                | Buzura suppressaria nuclear polyhedrosis virus gene for polyhedrin protein                           | 0.57    | <NONE>   | <NONE>      | <NONE>  |
| 381    | AJ012159                              | Homo sapiens 5T4 oncofetal trophoblast glycoprotein gene   | 0.57    | <NONE>   | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 382    | X76571                                | H.sapiens simple DNA sequence region clone wgl1a8.  | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 383    | AF034434                              | Vibrio cholerae pathogenicity island, putative transposase, aldehyde dehydrogenase (aldA), toxR-activated gene A protein (tagA), putative inner membrane protein, and putative zinc metalloprotease genes, complete cds; and... | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 384    | AB017031                              | Mus musculus gene for TESP4, complete cds   | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 385    | X89788                                | S.hispidus mitochondrial DNA for SSU ribosomal RNA gene   | 0.57    | <NONE>   | <NONE>  | <NONE>  |
| 386    | L16921                                | Rat progesteron receptor gene, 5' untranslated region.  | 0.57    | 3323116  | (AE001251) femA protein, putative [Treponema pallidum]                                      | 8.9     |
| 387    | AF027292                              | Homo sapiens interferon regulatory factor 6   | 0.57    | 259790   | (S48157) DNA polymerase-primase 180 kda subunit [Drosophila melanogaster, Peptide, 1490 aa] | 6.7     |
| 388    | AJ012581                              | Cicer arietinum mRNA for cytochrome P450  | 0.57    | 2131498  | hypothetical protein YDR446w - yeast CAI: 0.11 [Saccharomyces cerevisiae]                   | 5.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 389    | L15363                                | Human transfer RNA-Met (TRMEP1) pseudogene, complete gene                                       | 0.57    | 3228680  | (AF070935) GABA receptor subunit [Musca domestica]  | 5.2     |
| 390    | AE000525                              | Helicobacter pylori 26695 section 3 of 134 of the complete genome                               | 0.57    | 1938478  | (U97008) weak similarity to family 1 of G-protein coupled receptors [Caenorhabditis elegans]  | 4.0     |
| 391    | AF020189                              | Amblyomma americanum ecdysteroid receptor (AamEcR) mRNA, 3'UTR, region 1                        | 0.57    | 2072224  | (U94875) p40 [Borna disease virus]  | 4.0     |
| 392    | X56997                                | Human UbA52 gene coding for ubiquitin-52 amino acid fusion protein                              | 0.57    | 2960113  | (AL022121) hypothetical protein Rv3689  | 4.0     |
| 393    | AL010260                              | Plasmodium falciparum DNA ***<br>SEQUENCING IN PROGRESS *** from contig 4-81, complete sequence | 0.57    | 117233   | CYTOCHROME P450 2C14 (CYP11C14) phenobarbital-inducible, hepatic - rabbit P-450 [Oryctolagus cuniculus]<br>>gi 358265 prf  1306317A cytochrome P450 [Oryctolagus cuniculus] | 3.9     |
| 394    | M99581                                | Xenopus laevis gamma-crystallin (gcry3) gene, complete cds.                                     | 0.57    | 141647   | GASTRULA ZINC FINGER PROTEIN XLCGF44.2<br>>gi 85736 pir  S06571 finger protein (clone XlCGF44-2) - African clawed frog (fragment)   | 3.0     |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|--|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 395       | M38384                                | Drosophila melanogaster seven in absentia mRNA, complete cds.                    | 0.57    | 1707127   | (U80454) T16A1.1 [Caenorhabditis elegans]   | 3.0     |
| 396       | U32795                                | Haemophilus influenzae Rd section 110 of 163 of the complete genome              | 0.57    | 1173433   | IRON(III)-TRANSPORT SYSTEM PERMEASE PROTEIN SFUB >gi 152861 (M33815) protein (sufB)                 | 2.3     |
| 397       | X12600                                | Klebsiella pneumoniae nifX, nifU, nifS, nifV and nifW genes                      | 0.57    | 2909562   | (AL021925) hypothetical protein Rv2256c   | 1.4     |
| 398       | AB014526                              | Homo sapiens mRNA for KIAA0626 protein, complete cds                             | 0.57    | 482390  | insect-stage-specific protein - Trypanosoma cruzi >gi 162099 (M65021) insect stage-specific antigen | 0.61    |
| 399       | AF063587                              | Rhodococcus fascians strain NRRL-B-15096 hypothetical protein gene, complete cds | 0.57    | 4104321   | (AF034582) vesicle associated protein [Rattus norvegicus]   | 0.46    |
| 400       | L11117                                | Guinea pig estrone sulfotransferase gene.  | 0.57    | 82584   | alpha/beta-gliadin precursor (clone A212) - wheat   | 0.35    |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|---|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 401                                   | V00829    | Mouse complete gene for a mouse kallikrein gene. Genes are mGK 1 (complete gene) and mGK-2 of hormones, e.g., grow... > :: gb J00390 MUS KAL07 Mouse pseudo-kallikrein 2, exons 4 and 5, and kallikrein 1 gene, complete cds. | 0.57    | 2500916   | NUCLEAR HORMONE RECEPTOR NOR-2 receptor [Rattus norvegicus] >gi 1583604 prf 2121281A NOR-2 protein [Rattus norvegicus]  | 0.20    |
| 402                                   | X53092    | Chicken mRNA for beta-2 subunit of neuronal nicotinic acetylcholine receptor  | 0.57    | 1072256   | (U40953) similar to matrin F/G (SP:Q00910) containing C4-type zinc-fingers [Caenorhabditis elegans]   | 0.031   |
| 403                                   | L07939    | Ovis ovis granulocyte colony stimulating factor   | 0.57    | 3874345   | (Z81035) predicted using Genefinder; Similarity to dehydrogenases; cDNA EST EMBL:D65800 comes from this gene; cDNA EST EMBL:D76184 comes from this gene; cDNA EST EMBL:D69322 comes from this gene; cDNA EST EMBL:C08158 comes f... | 3e-07   |
| 404                                   | U18061    | Colletotrichum gloeosporioides CAP20 (cap20) gene, complete cds.  | 0.57    | 2914695   | (AC003974) putative ubiquitin specific protease   | 9e-08   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 405    | Z73955                                | L.japonicus mRNA for small GTP-binding protein, RAB11G  | 0.57    | 112894   | TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 3 (PUTATIVE DNA BINDING PROTEIN A20) (ZINC FINGER PROTEIN A20) >gi 107549 pir  A35797 probable DNA-binding protein A20 - human >gi 177866 (M59465) A20                                 | 7e-08   |
| 406    | X04335                                | Petunia grp-1 gene for glycine-rich protein   | 0.57    | 3876901  | (Z77660) Similarity to Human enoyl-CoA hydratase (SW:ECHM_HUMAN); cDNA EST EMBL:T00611 comes from this gene; cDNA EST yk203d10.3 comes from this gene; cDNA EST yk203d10.5 comes from this gene; cDNA EST yk457h5.3 comes from t... | 1e-27   |
| 407    | U40718                                | Rattus norvegicus S-adenosylmethionine decarboxylase (AMDP2) pseudogene                                   | 0.56    | <NONE>   | <NONE>  | <NONE>  |
| 408    | M60318                                | S.cerevisiae SSD1 protein gene, complete cds. > :: gb AR013983 AR013983 Sequence 8 from patent US 5773245 | 0.56    | <NONE>   | <NONE>  | <NONE>  |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|-----------|---------------------------------------|--|---------|---|-------------|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 409       | X60057                                | Nicotiana<br>tabacum blp4<br>mRNA for<br>luminal binding<br>protein (BiP)  | 0.56    | <NONE>  | <NONE>      | <NONE>  |
| 410       | AF085930                              | Homo sapiens<br>full length insert<br>cDNA clone<br>YR55A09  | 0.56    | <NONE>  | <NONE>      | <NONE>  |
| 411       | AL010189                              | Plasmodium<br>falciparum DNA<br>***<br>SEQUENCING<br>IN PROGRESS<br>*** from contig<br>3-102, complete<br>sequence | 0.56    | <NONE>  | <NONE>      | <NONE>  |
| 412       | X05402                                | Murine G-CSF<br>gene for<br>granulocyte<br>colony<br>stimulating<br>factor precursor                               | 0.56    | <NONE>  | <NONE>      | <NONE>  |
| 413       | U92280                                | Rattus<br>norvegicus<br>regulator of G-<br>protein<br>signalling 12<br>(RGS12)<br>mRNA,<br>complete cds            | 0.56    | <NONE>  | <NONE>      | <NONE>  |
| 414       | U85660                                | Human<br>papillomavirus<br>strain RTRX7<br>complete<br>genome  | 0.56    | <NONE>  | <NONE>      | <NONE>  |
| 415       | X57626                                | M. javanica<br>mitochondrion<br>ATPase 6, and<br>putative tRNA-f-<br>Met and tRNA-<br>His genes                    | 0.56    | <NONE>  | <NONE>      | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 416                                   | AB003363  | Sus scrofa<br>S100C gene,<br>complete cds  | 0.56    | <NONE>  | <NONE>  | <NONE>  |
| 417                                   | L42291    | Danio rerio<br>DANA element,<br>intron 4.  | 0.56    | 2650002   | (AE001062)<br>conserved hypothetical<br>protein<br>[Archaeoglobus<br>fulgidus]  | 8.7     |
| 418                                   | AF031826  | Mus musculus<br>leukocystatin<br>gene, complete<br>cds   | 0.56    | 462493  | L-LACTATE<br>DEHYDROGENASE<br>(IMMUNOGENIC<br>PROTEIN P36)<br>>gi 479296 pir  S3336<br>2 L-lactate<br>dehydrogenase (EC<br>1.1.1.27) -<br>Mycoplasma<br>hyopneumoniae   | 6.7     |
| 419                                   | U17068    | Pennisetum<br>glaucum Ac-like<br>element, AcL2.  | 0.56    | 399449  | ESCARGOT/SNAIL<br>PROTEIN<br>HOMOLOG  | 6.7     |
| 420                                   | Z48042    | H.sapiens<br>mRNA encoding<br>GPI-anchored<br>protein p137   | 0.56    | 141232  | HYPOTHETICAL 8.7<br>KD PROTEIN<br>(READING FRAME<br>D)<br>>gi 76316 pir  QQSA7<br>Chypothetical protein<br>E-74   | 6.7     |
| 421                                   | AF027657  | Choristoneura<br>fumiferana<br>entomopoxvirus<br>nucleotide<br>triphosphate<br>phosphohydrolas<br>e I (NPHI) gene,<br>complete cds | 0.56    | 464999  | PUTATIVE<br>ACETYLCHOLINE<br>REGULATOR UNC-<br>18<br>>gi 480359 pir  S3674<br>7 acetylcholine<br>regulator unc-18 -<br>Caenorhabditis<br>elegans<br>>gi 247392 bbs 10029<br>4 putative<br>acetylcholine regulator<br>unc-18 | 5.1     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 422    | AB011540                              | Homo sapiens mRNA for MEGF7, partial cds  | 0.56    | 1718033  | URACIL-DNA GLYCOSYLASE (UDG) herpesvirus 2 >gi 695219 (U20824) uracil DNA glycosylase  | 5.1     |
| 423    | X59941                                | X.maculatus NGF gene for nerve growth factor                                      | 0.56    | 1169081  | COMMON PLANT REGULATORY FACTOR CPRF-1 >gi 515621 (X58575) light-inducible protein CPRF-1 [Petroselinum crispum] >gi 1498301 (U46217) CPRF1   | 3.8     |
| 424    | M72711                                | Rat transcriptional repressor of myelin-specific genes (SCIP) mRNA, complete cds. | 0.56    | 501027   | (U01849) ORF2 [Trypanosoma brucei]   | 2.3     |
| 425    | AL023850                              | Caenorhabditis elegans cosmid Y67D11A, complete sequence [Caenorhabditis elegans] | 0.56    | 266771   | CHORISMATE MUTASE (CM) / PREPHENATE DEHYDRATASE (PDT) (P-PROTEIN) >gi 281791 pir  S2605 3 chorismate mutase (EC 5.4.99.5) P / prephenate dehydratase (EC 4.2.1.51) - Erwinia herbicola >gi 43344 | 2.3     |
| 426    | U47862                                | Schistosoma mansoni gynecophoral canal protein mRNA, complete cds                 | 0.56    | 2147138  | ATP synthase chain 6 - Platymonas subcordiformis mitochondrion >gi 633582 (Z47797) ATP synthase subunit 6 [Platymonas subcordiformis]  | 2.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 427    | V00574                                | Human germ line gene homologous to bladder carcinoma oncogene T24 (Gene code c-Ha-ras-1) with four exons. | 0.56    | 1518672  | (U60289) receptor protein tyrosine phosphatase psi [Homo sapiens] | 1.7     |
| 428    | Z71502                                | X.laevis H1(0)-1 gene   | 0.56    | 1651674  | (D90899) ferrichrome-iron receptor                                | 1.3     |
| 429    | M37278                                | R.norvegicus renin gene, exons 1-9.   | 0.56    | 2853019  | (AF045141) putative serine proteinase [Scirpophaga incertulas]    | 1.0     |
| 430    | D28878                                | Thermus thermophilus polA gene for thermostable DNA polymerase I, complete cds                            | 0.56    | 3659692  | (AF068748) sphingosine kinase [Mus musculus]                      | 0.77    |
| 431    | Z15027                                | H.sapiens HLA class III DNA   | 0.56    | 1304141  | (D43758) fibrinogen A-alpha-chain                                 | 0.76    |
| 432    | M14362                                | Human T-cell surface antigen CD2 (T11) mRNA, complete cds.  | 0.56    | 2462979  | (Y11915) Tenascin-X [Bos taurus]                                  | 0.59    |
| 433    | Z50801                                | Z.mays mRNA for chlorophyll a/b-binding protein CP29  | 0.56    | 109677   | collagen alpha 1(I) chain - mouse >gi 50487                       | 0.50    |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 434       | Z38114                                | S.cerevisiae<br>chromosome<br>XIII cosmid<br>9745  | 0.56    | 140372  | HYPOTHETICAL<br>86.0 KD PROTEIN IN<br>GLK1-SRO9<br>INTERGENIC<br>REGION<br>>gi 83159 pir  S19367<br>hypothetical protein<br>YCL039w - yeast<br>(Saccharomyces<br>cerevisiae) | 0.35    |
| 435       | AF052254                              | Escherichia coli<br>DNA gyrase A<br>(gyrA) gene,<br>partial cds  | 0.56    | 2724126   | (AF038535)<br>synaptotagmin VII<br>[Homo sapiens]  | 0.12    |
| 436       | AF080649                              | Tegula pulligo<br>12S small<br>subunit<br>ribosomal RNA<br>gene,<br>mitochondrial<br>gene for<br>mitochondrial<br>RNA, partial<br>sequence | 0.56    | 3913223   | CYCLIN-<br>DEPENDENT<br>KINASE INHIBITOR<br>1 p21/WAF1 [Felis<br>catus]  | 0.11    |
| 437       | AJ005690                              | Danio rerio<br>mRNA for<br>protein tyrosine<br>kinase  | 0.56    | 2623830   | (AF030962) unknown<br>[Schistosoma<br>mansoni]   | 7e-06   |
| 438       | U31202                                | Human noggin<br>(NOGGIN)<br>gene, complete<br>cds.   | 0.56    | 3875475   | (Z78411) F02D8.3<br>[Caenorhabditis<br>elegans]  | 1e-06   |
| 439       | X51695                                | Ovis sp.<br>trichohyalin<br>mRNA, partial  | 0.56    | 3386622   | (AC004665) unknown<br>protein [Arabidopsis<br>thaliana]  | 1e-10   |
| 440       | U28938                                | Rattus<br>norvegicus<br>protein tyrosine<br>phosphatase D30<br>mRNA,<br>complete cds   | 0.56    | 3293547   | (AF072709) putative<br>oxidoreductase<br>[Streptomyces<br>lividans]  | 1e-14   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 441    | AE001171                              | Borrelia burgdorferi (section 57 of 70) of the complete genome  | 0.56    | 2315521  | (AF016452) similar to the beta transducin family   | 4e-16   |
| 442    | AF036685                              | Caenorhabditis elegans cosmid C05B10                            | 0.56    | 1519671  | (U67951) contains similarity to ATP/GTP-binding site motif (PS:PS00017) [Caenorhabditis elegans] | 6e-20   |
| 443    | X01173                                | Xenopus laevis vitellogenin gene A1 5' flanking region          | 0.56    | 1118102  | (U41558) K02B2.3 gene product [Caenorhabditis elegans]   | 2e-31   |
| 444    | D10911                                | Mus musculus DNA for MS2 protein, complete cds                  | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 445    | D30010                                | Rice mRNA EN117, partial sequence                               | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 446    | U51991                                | Escherichia coli phosphoprotein phosphatase                     | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 447    | M18858                                | Mouse T cell receptor C-gamma-7.1 mRNA, 3' end.                 | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 448    | U95218                                | Homo sapiens T cell-death associated protein gene, complete cds | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 449    | M14948                                | Human R-ras gene, exon 1.                                       | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 450    | AB002353                              | Human mRNA for KIAA0355 gene, complete cds                      | 0.55    | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 451    | L81689                                | Homo sapiens (subclone 1_d6 from P1 H54) DNA sequence   | 0.55    | <NONE>   | <NONE>   | <NONE>  |
| 452    | M68955                                | Human myristoylated alanine-rich C-kinase substrate (MACS) gene, 5' end.  | 0.55    | 3322710  | (AE001220) V-type ATPase, subunit B (atpB-1) [Treponema pallidum]  | 5.0     |
| 453    | X62953                                | R.norvegicus mRNA (pJG116) with repetitive elements   | 0.55    | 1076802  | extensin-like protein - maize >gi 600118 mays]   | 5.0     |
| 454    | L34630                                | Synechocystis sp. mntABC transporter system: periplasmic-binding protein (mntC), complete cds; (mntA) gene, complete cds; membrane protein (mntB) gene, complete cds. | 0.55    | 2117632  | hydrogen dehydrogenase (EC 1.12.1.2) - Clostridium acetobutylicum >gi 557064 (U15277) hydrogenase I [Clostridium acetobutylicum] | 5.0     |
| 455    | U43521                                | Plasmodium berghei merozoite surface protein-1 gene, complete cds   | 0.55    | 127654   | MYOGLOBIN  | 4.9     |
| 456    | Z64937                                | H.sapiens CpG DNA, clone 17g7, reverse read cpg 17g7.rtl.   | 0.55    | 417298   | MFS18 PROTEIN PRECURSOR  | 3.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 457    | U10914                                | Macaca mulatta clone irh83 T-cell receptor alpha chain mRNA, partial cds.                   | 0.55    | 310406   | (L09212) tat protein [Simian immunodeficiency virus] virus]   | 3.8     |
| 458    | AF022838                              | Homo sapiens multidrug resistance protein   | 0.55    | 1585251  | traB gene [Amycolatopsis methanolica]   | 2.8     |
| 459    | M35603                                | Mouse Hox-3.1 gene and Hox-3.2-Hox-3.1 intergenic region.                                   | 0.55    | 818849   | (U25430) nucleotide pyrophosphatase precursor [Oryza sativa]  | 2.0     |
| 460    | AE001395                              | Plasmodium falciparum chromosome 2, section 32 of 73 of the complete sequence               | 0.55    | 137532   | PROTEIN C2 >gi 74386 pir  WZVZB 6 59K HindIII-C protein - vaccinia virus (strain WR)  | 1.7     |
| 461    | AE001395                              | Plasmodium falciparum chromosome 2, section 32 of 73 of the complete sequence               | 0.55    | 137532   | PROTEIN C2 >gi 74386 pir  WZVZB 6 59K HindIII-C protein - vaccinia virus (strain WR)  | 1.7     |
| 462    | U59736                                | Human transcription factor (NFATc.b) mRNA, complete cds                                     | 0.55    | 3327144  | (AB014565) KIAA0665 protein [Homo sapiens]  | 0.096   |
| 463    | U34860                                | Saccharomyces cerevisiae origin recognition complex large subunit (ORC1) gene, complete cds | 0.55    | 140372   | HYPOTHETICAL 86.0 KD PROTEIN IN GLK1-SRO9 INTERGENIC REGION >gi 83159 pir  S19367 hypothetical protein YCL039w - yeast (Saccharomyces cerevisiae) | 0.017   |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 464                                   | AF012341  | Homo sapiens glutaryl-CoA dehydrogenase (GCDH) gene, exons 6, 7, 8, 9, and 10   | 0.55    | 1166611   | (U46674) coded for by C. elegans cDNA yk27d9.5; coded for by C. elegans cDNA yk27d9.3; short region of weak homology to drosophila suppressor of sable protein | 0.008   |
| 465                                   | AF004891  | HIV-1 isolate Q98-CxA from Kenya, envelope glycoprotein C2V3 region (env) gene, partial cds                           | 0.54    | <NONE>  | <NONE>   | <NONE>  |
| 466                                   | Y10159    | D.discoideum racGAP gene  | 0.54    | <NONE>  | <NONE>   | <NONE>  |
| 467                                   | AB001895  | Homo sapiens mRNA for B120, complete cds  | 0.54    | <NONE>  | <NONE>   | <NONE>  |
| 468                                   | X12357    | Bovine gene for aspartyl protease NM1 exons 3 and 4 > :: 1c1 X12357 Bovine aspartyl protease NM1 gene, exons 3 and 4. | 0.54    | <NONE>  | <NONE>   | <NONE>  |
| 469                                   | AE001151  | Borrelia burgdorferi (section 37 of 70) of the complete genome  | 0.54    | <NONE>  | <NONE>   | <NONE>  |
| 470                                   | X92052    | H.sapiens mRNA for T cell receptor alpha chain  | 0.54    | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 471    | U00938                                | Mus musculus ileal lipid-binding protein gene, complete cds  | 0.54    | 1009712  | (U27698) calreticulin [Arabidopsis thaliana]  | 4.9     |
| 472    | X68367                                | M.thermoformicum complete plasmid pFZ1 DNA   | 0.54    | 125272   | CASEIN KINASE II, ALPHA CHAIN (CK II)<br>>gi 419938 pir  A43297 casein kinase II (EC 2.7.1.-) alpha chain - Theileria parva<br>>gi 161871 (M92084) casein kinase II alpha subunit [Theileria parva] | 4.7     |
| 473    | Z61098                                | H.sapiens CpG DNA, clone 44c4, reverse read cpg44c4.rtl.a.   | 0.54    | 4191274  | (AJ131094) Xvent-1B protein [Xenopus laevis]  | 3.7     |
| 474    | M63962                                | Human gastric H,K-ATPase catalytic subunit gene, complete cds.   | 0.54    | 3881648  | (Z70757) similar to serine protease inhibitor [Caenorhabditis elegans]  | 3.7     |
| 475    | X86019                                | H.sapiens mRNA for PRPL-2 protein  | 0.54    | 1648828  | (D87963) ETF-related factor-1 (ETFR-1)  | 2.1     |
| 476    | X89010                                | S.glaucescens genes strU, strX, strV and strW for 5'-hydroxystreptomycin production and transport polypeptides | 0.54    | 3550345  | (AF084524) cellular repressor of E1A-stimulated genes CREG [Mus musculus]   | 0.25    |
| 477    | AB007836                              | Homo sapiens mRNA for Hic-5, partial cds   | 0.54    | 1097213  | ORF 1 [Streptomyces lavendulae]   | 0.15    |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION                                | P VALUE |
| 478    | U32622                                | Comamonas testosteroni TsaR (tsaR), toluenesulfonate methyl-monooxygenase oxygenase component component (tsaB), toluenesulfonate zinc-independent alcohol dehydrogenase ... | 0.54    | 3875351  | (Z96047) DY3.6 [Caenorhabditis elegans]    | 0.006   |
| 479    | D61394                                | Arabidopsis thaliana gene for beta-VPE, complete cds  | 0.53    | <NONE>   | <NONE>                                     | <NONE>  |
| 480    | D61394                                | Arabidopsis thaliana gene for beta-VPE, complete cds  | 0.53    | <NONE>   | <NONE>                                     | <NONE>  |
| 481    | Z33072                                | M.capricolum DNA for CONTIG MC097   | 0.53    | <NONE>   | <NONE>                                     | <NONE>  |
| 482    | U45975                                | Human phosphatidylinositol (4,5)bisphosphate 5-phosphatase homolog mRNA, partial cds.   | 0.53    | <NONE>   | <NONE>                                     | <NONE>  |
| 483    | Z71324                                | S.cerevisiae chromosome XIV reading frame ORF YNL048w   | 0.53    | 2135586  | M130 antigen (cytosolic variant 1) - human | 2.1     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 484                                   | L32090    | Listeria monocytogenes secA gene, complete cds.  | 0.53    | 2291129   | (AF016415) No definition line found [Caenorhabditis elegans]               | 0.70    |
| 485                                   | D86423    | Mus musculus mRNA for HGT keratin, partial cds   | 0.53    | 1235974   | (X96713) collagen [Globodera pallida]                                      | 0.41    |
| 486                                   | Y15969    | Mus musculus V kappa 21-6 gene, partial  | 0.52    | <NONE>  | <NONE>   | <NONE>  |
| 487                                   | M27480    | Mus musculus (clone 3F9) transcribed germline T cell receptor gamma chain (Tcr-g) mRNA, VJ4 C4 region. | 0.52    | 3875542   | (Z67990) Similarity to Rat amiloride-sensitive sodium channel beta-subunit | 4.6     |
| 488                                   | D87004    | Human (lambda) DNA for immunoglobulin light chain  | 0.52    | 1766073   | (U37272) winged helix protein CWH-1 [Gallus gallus]                        | 3.5     |
| 489                                   | Z99704    | Human DNA sequence from cosmid E75B8 on chromosome 22, complete sequence [Homo sapiens]                | 0.51    | <NONE>  | <NONE>   | <NONE>  |
| 490                                   | U76523    | Sambucus nigra lectin precursor mRNA, complete cds   | 0.51    | <NONE>  | <NONE>   | <NONE>  |
| 491                                   | U32795    | Haemophilus influenzae Rd section 110 of 163 of the complete genome                                    | 0.50    | <NONE>  | <NONE>   | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 492                                   | M14602    | Human<br>myoglobin gene,<br>exon 2.  | 0.49    | 478384  | helicase homolog<br>g10L protein - African<br>swine fever virus<br>>gi414091 (X72951)<br>G10L 125 KDa protein   | 7.0     |
| 493                                   | D87075    | Human mRNA<br>for KIAA0238<br>gene, partial cds  | 0.24    | 1938429   | (U97002) similar to<br>Schizosaccharomyces<br>pombe 4-<br>nitrophenylphosphatase<br>(PNPPASE)<br>(SP:Q00472,<br>NID:g5004)<br>[Caenorhabditis<br>elegans] | 2.5     |
| 494                                   | U95102    | Xenopus laevis<br>mitotic<br>phosphoprotein<br>90 mRNA,<br>complete cds                  | 0.23    | <NONE>  | <NONE>  | <NONE>  |
| 495                                   | J05254    | N.crassa<br>mitochondrial<br>small (19S)<br>rRNA and Cys-<br>tRNA.                       | 0.23    | 192150  | (L05670) clustrin<br>[Mus musculus]   | 5.1     |
| 496                                   | X16399    | Gene for<br>glutamate<br>dehydrogenase<br>(EC 1.4.1.4),<br>put. bacterial<br>origin      | 0.23    | 790933  | (L07867) invariant<br>surface glycoprotein<br>[Trypanosoma brucei]  | 0.030   |
| 497                                   | AE001251  | Treponema<br>pallidum section<br>67 of 87 of the<br>complete<br>genome                   | 0.22    | <NONE>  | <NONE>  | <NONE>  |
| 498                                   | AF026919  | Homo sapiens<br>amyloid lambda<br>light chain<br>variable region<br>mRNA, partial<br>cds | 0.21    | <NONE>  | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 499    | Z27247                                | D.melanogaster mRNA for defensin   | 0.21    | <NONE>   | <NONE>  | <NONE>  |
| 500    | Y15608                                | Candida albicans UBI3 gene   | 0.21    | <NONE>   | <NONE>  | <NONE>  |
| 501    | V00598                                | Human beta-tubulin pseudogene.   | 0.21    | <NONE>   | <NONE>  | <NONE>  |
| 502    | X79426                                | A.thaliana microsatellite [repeated motif (gat)7]                            | 0.21    | <NONE>   | <NONE>  | <NONE>  |
| 503    | X75772                                | A.caerulescens mitochondrial genes for cytochrome b and NADH dehydrogenase 5 | 0.21    | 139626   | PROTEIN T1 PRECURSOR  | 7.8     |
| 504    | AF028736                              | Serratia marcescens site specific recombinase                                | 0.21    | 3645960  | (AL031583) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=47.46; 1-evidence_end; 2-evidence=predicted by match; 2-match_accession=SWISS-PROT:P23792; 2-match_description=DISCONNECTED PROTEIN.; 2-matc... | 4.6     |
| 505    | X97545                                | S.cerevisiae OST5 gene   | 0.21    | 2275631  | (AF014940) No definition line found [Caenorhabditis elegans]  | 2.7     |
| 506    | M24543                                | Human prostate-specific antigen (PA) gene, complete cds.                     | 0.21    | 1938527  | (U97012) C04E6.2 gene product [Caenorhabditis elegans]  | 2.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 507    | M62470                                | Mouse thrombospondin (THBS1) gene, complete cds.          | 0.21    | 548563   | RNA REPLICASE POLYPROTEIN 2.7.7.48) - Erysimum latent virus >gi 3892232 (AF098523) replicase protein [Erysimum latent virus]               | 2.1     |
| 508    | Y13544                                | Homo sapiens cosmid C1                                    | 0.21    | 1235710  | (L40584) polypeptide [Infectious pancreatic necrosis virus]  | 2.0     |
| 509    | M24193                                | Chicken MHC B complex protein (C12-3) mRNA, complete cds. | 0.21    | 3600102  | (AF090441) extracellular reelin [Gallus gallus]  | 0.52    |
| 510    | X97161                                | H.sapiens TFE3 gene, exon 4,5 & 6                         | 0.21    | 854065   | (X83413) U88 [Human herpesvirus 6]   | 0.30    |
| 511    | X67649                                | R.norvegicus DNA sequence for LFB1/HNF1 promoter          | 0.21    | 3913114  | TRANSCRIPTION FACTOR COUP 2 COUP-TFII - chicken >gi 392817 (U00697) orphan receptor COUP-TFII [Gallus gallus]                              | 0.004   |
| 512    | U63807                                | Fugu rubripes growth hormone (GH) gene, complete cds      | 0.21    | 3510505  | (AF030881) pol polyprotein [Fugu rubripes]   | 3e-04   |
| 513    | Z95636                                | H.sapiens mRNA for laminin alpha 5 chain                  | 0.21    | 400350   | NAM7 PROTEIN (NONSENSE MEDIATED MRNA DECAY PROTEIN 1) (UP-FRAMESHIFT SUPPRESSOR 1) factor NAM7 - yeast (Saccharomyces cerevisiae) >gi 4023 | 1e-07   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|---|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 514    | U91907                                | Mirounga leonina major histocompatibility complex class II (DQA) gene, partial cds                    | 0.20    | <NONE>   | <NONE>      | <NONE>  |
| 515    | Z35758                                | Transmissible gastroenteritis virus TFI virion protein genes  | 0.20    | <NONE>   | <NONE>      | <NONE>  |
| 516    | X00334                                | Drosophila virilis simple DNA sequence (pDv-19)   | 0.20    | <NONE>   | <NONE>      | <NONE>  |
| 517    | M76741                                | Homo sapiens biliary glycoprotein (BGP) gene, partial cds.  | 0.20    | <NONE>   | <NONE>      | <NONE>  |
| 518    | D78515                                | Mus musculus rae28 gene, exon 1 and 5'flanking region   | 0.20    | <NONE>   | <NONE>      | <NONE>  |
| 519    | M62975                                | Drosophila melanogaster RNA polymerase II second largest subunit upstream (DmRP 140) gene, exons 1-4. | 0.20    | <NONE>   | <NONE>      | <NONE>  |
| 520    | M27260                                | Chicken 78-kD glucose-regulated protein, complete cds.  | 0.20    | <NONE>   | <NONE>      | <NONE>  |
| 521    | AF076470                              | Rice tungro bacilliform virus Serdang strain, complete genome   | 0.20    | <NONE>   | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 522    | AF076470                              | Rice tungro bacilliform virus Serdang strain, complete genome                                   | 0.20    | <NONE>   | <NONE>  | <NONE>  |
| 523    | U04636                                | Human cyclooxygenase-2 (hCox-2) gene, complete cds.   | 0.20    | <NONE>   | <NONE>  | <NONE>  |
| 524    | AE001430                              | Plasmodium falciparum chromosome 2, section 67 of 73 of the complete sequence                   | 0.20    | <NONE>   | <NONE>  | <NONE>  |
| 525    | AF043514                              | Mus musculus phosphomannomutase (Pmm2) mRNA, complete cds                                       | 0.20    | 3025006  | HYPOTHETICAL 15.5 KD PROTEIN IN MOAE-RHLE INTERGENIC REGION >gi 1787009 (AE000181) orf, hypothetical protein [Escherichia coli] | 9.8     |
| 526    | U23144                                | Xenopus laevis FTZ-F1-related nuclear orphan receptor variant (xFF1rAshort) mRNA, complete cds. | 0.20    | 3184402  | (AB014477) period protein [Chymomyza costata]   | 9.6     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 527    | U14621                                | Paracentrotus lividus Pax-6 (suPax-6) mRNA, complete cds.  | 0.20    | 465894   | PROBABLE MICROSOMAL SIGNAL PEPTIDASE 23 KD SUBUNIT (SPC22/23) >gi 630688 pir  S4485 4 K12H4.4 protein - Caenorhabditis elegans >gi 289708 (L14331) homology with signal peptidase; coded for by C. elegans cDNAs GenBank: M79661, M79662 and M79663; putative | 7.7     |
| 528    | AF030511                              | Actinobacillus pleuropneumoniae MRP ATPase homolog (mrp) gene, partial cds; ApxIVA var3 (apxIVA) gene, complete cds; and beta-galactosidase (lacZ) gene, partial cds | 0.20    | 1175966  | HYPOTHETICAL 45.1 KD PROTEIN IN THE 5'REGION >gi 1084720 pir  S5619 3 probable membrane protein YFL062w - yeast (Saccharomyces cerevisiae)  | 7.2     |
| 529    | AF070581                              | Homo sapiens clone 24540 mRNA sequence   | 0.20    | 542394   | glyoxal oxidase (EC 1.2.3.-) precursor - basidiomycete (Phanerochaete chrysosporium) >gi 1050302  | 5.8     |
| 530    | X75437                                | T.maritima pgK gene for 3-phosphoglycerate kinase  | 0.20    | 825648   | (Z34531) coproporphyrinogen oxidase [Homo sapiens]  | 5.8     |
| 531    | U32686                                | Haemophilus influenzae Rd section 1 of 163 of the complete genome  | 0.20    | 3309593  | (AF072878) ciliary outer arm dynein beta heavy chain  | 5.6     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 532    | Z28081                                | S.cerevisiae chromosome XI reading frame ORF YKL081w         | 0.20    | 2507201  | CARBON CATABOLITE DEREPRESSING PROTEIN KINASE >gi 1469803 (L78129) serine/threonine kinase [Candida albicans]   | 5.5     |
| 533    | AF022725                              | Hordeum vulgare limit dextrinase (HvLD99) gene, complete cds | 0.20    | 3139154  | (AF064077) adrenocorticotrophic hormone receptor [Sus scrofa]   | 4.3     |
| 534    | AL021726                              | Drosophila melanogaster cosmid 171E4                         | 0.20    | 3885334  | (AC005623) putative argonaute protein [Arabidopsis thaliana]  | 2.6     |
| 535    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds                   | 0.20    | 4008334  | (Z92824) B0413.4 [Caenorhabditis elegans]   | 1.5     |
| 536    | Z46606                                | H.sapiens HLTF gene for helicase-like transcription factor   | 0.20    | 132946   | 60S RIBOSOMAL PROTEIN L30B (RP29) cytosolic - yeast (Saccharomyces cerevisiae) >gi 171821 not determined) [Saccharomyces cerevisiae] >gi 1045254 cerevisiae] >gi 1323250 gnl PID e 243708 (Z72933) ORF YGR148c [Saccharomyces cerevisiae] | 1.5     |
| 537    | X87193                                | H.sapiens mRNA for 2.19 gene                                 | 0.20    | 139820   | DNA-REPAIR PROTEIN XRCC1  | 1.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 538    | L77965                                | Clostridium perfringens C beta 2 toxin gene, complete cds    | 0.20    | 1175950  | HYPOTHETICAL 33.5 KD PROTEIN IN SEC53-ACT1 INTERGENIC REGION<br>>gi 1084703 pir  S56211 probable membrane protein YFL044c - yeast (Saccharomyces cerevisiae)<br>>gi 836711 gnl PID d1009835 (D50617) YFL044C | 1.4     |
| 539    | M15938                                | Chicken neural cell-adhesion molecule (N-CAM) gene, exon 18. | 0.20    | 2133082  | regulatory protein MSR1 - yeast  | 1.1     |
| 540    | AJ003220                              | Solanum tuberosum mRNA for extensin-like protein, partial    | 0.20    | 2496932  | HYPOTHETICAL 55.9 KD PROTEIN C56G2.1 IN CHROMOSOME III<br>>gi 726413 (U23177) C56G2.1 gene product [Caenorhabditis elegans]  | 1.1     |
| 541    | X98108                                | A.thaliana psbP gene   | 0.20    | 119227   | EPIDERMAL GROWTH FACTOR PRECURSOR precursor - mouse<br>>gi 309210 (J00380) prepro-egf [Mus musculus]   | 0.49    |
| 542    | AB011179                              | Homo sapiens mRNA for KIAA0607 protein, partial cds          | 0.20    | 2143753  | gene VGF protein - rat<br>>gi 205690 (M60525) nerve growth factor inducible protein [Rattus norvegicus]<br>>gi 205701 (M60522) nerve growth factor-inducible protein [Rattus norvegicus]<br>>gi 207651       | 0.39    |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 543       | X75318                                | H.sapiens ITIH1<br>gene (exon 22)<br>and ITIH3 gene  | 0.20    | 629557  | RNA-binding protein<br>rnpD - Arabidopsis<br>thaliana (fragment)<br>>gi 510240 (X61108)<br>RNA binding protein<br>[Arabidopsis thaliana] | 0.38    |
| 544       | AB008374                              | Oncorhynchus<br>mykiss mRNA<br>for alpha 3 type I<br>collagen, partial<br>cds  | 0.20    | 1082610   | muf1 protein - human<br>>gi 762953 (X86018)<br>muf1 [Homo sapiens]   | 0.37    |
| 545       | U09809                                | Limulus<br>polyphemus<br>arginine kinase<br>mRNA,<br>complete cds.   | 0.20    | 3882016   | (AJ012650) CP<br>[Papaya ringspot<br>virus]  | 0.37    |
| 546       | AB020671                              | Homo sapiens<br>mRNA for<br>KIAA0864<br>protein, partial<br>cds  | 0.20    | 2674350   | (U93121) M-phase<br>phosphoprotein-1<br>[Homo sapiens]   | 0.18    |
| 547       | L04457                                | Phytophthora<br>megasperma<br>mitochondrial<br>ORF152,<br>complete cds,<br>cytochrome c<br>oxidase subunit I<br>(cox1) gene,<br>complete cds,<br>cytochrome c<br>oxidase subunit<br>II | 0.20    | 746516  | (U23517) D1022.7<br>[Caenorhabditis<br>elegans] >gi 3258651<br>elegans]  | 0.043   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 548    | L04457                                | Phytophthora megasperma mitochondrial ORF152, complete cds, cytochrome c oxidase subunit I (cox1) gene, complete cds, cytochrome c oxidase subunit II | 0.20    | 746516   | (U23517) D1022.7 [Caenorhabditis elegans] >gi 3258651 elegans]                          | 0.042   |
| 549    | S82819                                | Cdk5r=cyclin-dependent kinase 5 regulatory subunit p35 [mice, brain, 129/SvJ, C57BL/6, Genomic/mRNA, 5528 nt]   | 0.20    | 3413870  | (AB007923) KIAA0454 protein [Homo sapiens]  | 0.020   |
| 550    | D31792                                | Streptomyces griseus DNA for serine/threonine protein kinases, complete cds   | 0.20    | 861405   | (U29154) T07F12.2 gene product [Caenorhabditis elegans]                                 | 0.019   |
| 551    | U97499                                | Homo sapiens butyrophilin (BT3.2) gene, exons 5-10, and complete cds  | 0.20    | 2773341  | (AF040954) putative protein phosphatase 1 nuclear targeting subunit [Rattus norvegicus] | 0.008   |
| 552    | U31463                                | Rattus norvegicus nonmuscle myosin heavy chain-A mRNA, complete cds.  | 0.20    | 3880111  | (Z81130) predicted using Genefinder   | 0.002   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 553    | X78401                                | Bacteriophage P22 right operon, orf 48, replication genes 18 and 12, nin region genes, ninG phosphatase, late control gene 23, orf 60, complete cds, late control region, start of lysis gene 13 | 0.20    | 1123087  | (U42436) C49H3.3 gene product [Caenorhabditis elegans]  | 4e-04   |
| 554    | X57310                                | Nocardia lactamdurans pcbAB and pcbC genes for alpha-aminoadipyl-L-cysteiny-D-valine synthetase and isopenicillin N synthase   | 0.20    | 1723511  | PUTATIVE ENDONUCLEASE C1F12.06C yeast (Schizosaccharomyces pombe) >gi 1217980 (Z69944) unknown [Schizosaccharomyces pombe]  | 4e-09   |
| 555    | X62386                                | S.epidermidis genes epiY', epiY, epiA, epiB, epiC, epiD, epiQ, epiP  | 0.20    | 3874927  | (Z73424) C44B9.1 [Caenorhabditis elegans]   | 3e-10   |
| 556    | X59000                                | Epizootic haemorrhagic disease virus gene segment 6 for NS1  | 0.20    | 3879755  | (Z80220) similar to nucleotide binding protein; cDNA EST EMBL:M75897 comes from this gene; cDNA EST EMBL:M89054 comes from this gene; cDNA EST EMBL:D26713 comes from this gene; cDNA EST EMBL:D26718 comes from this gene; cDNA... | 8e-16   |
| 557    | M98776                                | Human keratin 1 gene, complete cds   | 0.20    | 1086900  | (U41278) contains similarity to G beta repeats  | 2e-30   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|-----------|---------------------------------------|--|---------|---|-------------|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 558       | AF011446                              | Mus musculus<br>granzyme K<br>gene, complete<br>cds  | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 559       | AF074708                              | Macaca mulatta<br>clone MMU1.5<br>FRG1-like<br>pseudogene,<br>exons 7 and 8,<br>partial sequence | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 560       | X13287                                | Medicago sativa<br>nodulin-25 gene   | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 561       | Z49509                                | S.cerevisiae<br>chromosome X<br>reading frame<br>ORF YJR009c                                     | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 562       | D89041                                | Bovine DNA for<br>prostaglandin<br>F2alpha<br>receptor, partial<br>cds                           | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 563       | D29644                                | Streptococcus<br>salivarius DNA<br>for dextranase  | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 564       | AE001461                              | Helicobacter<br>pylori, strain J99<br>section 22 of 132<br>of the complete<br>genome             | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 565       | L38559                                | Homo sapiens<br>galactocerebrosi<br>dase (GALC)<br>gene, exon 17.                                | 0.19    | <NONE>  | <NONE>      | <NONE>  |
| 566       | Z82628                                | R.prowazekii<br>genomic DNA<br>fragment (clone<br>A405F)   | 0.19    | <NONE>  | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|---|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 567    | U25641                                | Tetrahymena thermophila telomerase component p80 mRNA, complete cds   | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 568    | AB002343                              | Human mRNA for KIAA0345 gene, complete cds  | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 569    | D10064                                | Erwinia carotovora gene for pectate lyase III, complete cds   | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 570    | U31734                                | Homo sapiens clone MF118 A4A10 hypoxanthine phosphoribosyltransferase (hprt) 130 kb deletion mutant mRNA, partial cds, contains human Alu element | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 571    | AE001386                              | Plasmodium falciparum chromosome 2, section 23 of 73 of the complete sequence   | 0.19    | <NONE>   | <NONE>      | <NONE>  |
| 572    | M95623                                | Homo sapiens hydroxymethylbilane synthase gene, complete cds.   | 0.19    | <NONE>   | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 573    | S67478                                | (GC*IS)=vitamin D-binding protein/group specific component [human, peripheral blood leukocytes, Genomic, 794 nt, segment 4 of 9] | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 574    | X99075                                | H.sapiens NRGN gene, exon 1  | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 575    | AF044775                              | Homo sapiens breakpoint cluster region BCRder14 sequence   | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 576    | AB002333                              | Human mRNA for KIAA0335 gene, complete cds   | 0.19    | <NONE>   | <NONE>  | <NONE>  |
| 577    | U53566                                | Macaca mulatta pit-1/GHF-1 transcription factor mRNA, complete cds   | 0.19    | 1078068  | probable membrane protein YLR311c - yeast   | 9.2     |
| 578    | U73664                                | Human t(11;14)(q13;q32) breakpoint junction sequence   | 0.19    | 116734   | COAT PROTEIN (CAPSID PROTEIN) virus >gi 58901 (X62133) CyMV coat protein gene product | 8.8     |
| 579    | AF004054                              | Heterophyllaea pustulata rps 16 gene, chloroplast gene, partial intron sequence  | 0.19    | 1928991  | (U92815) heat shock protein 70 precursor [Citrullus lanatus]                          | 8.7     |



| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|---|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 580       | X27081                                | Caenorhabditis<br>elegans cosmid<br>M01A8,<br>complete<br>sequence<br>[Caenorhabditis<br>elegans]   | 0.19    | 2496247   | HYPOTHETICAL<br>ATP-BINDING<br>PROTEIN MJ0625<br>>gi 2128413 pir  A643<br>78 hypothetical protein<br>MJ0625 -<br>Methanococcus<br>jannaschii<br>>gi 1591336 (U67510)<br>M. jannaschii<br>predicted coding<br>region MJ0625 | 8.6     |
| 581       | Z74145                                | S.cerevisiae<br>chromosome IV<br>reading frame<br>ORF YDL097c   | 0.19    | 1174425   | TYROSINE-<br>PROTEIN KINASE<br>SPK-1   | 6.7     |
| 582       | D38547                                | Small round<br>structured virus<br>genomic RNA,<br>3'terminal<br>sequence<br>containing ORF2<br>and ORF3  | 0.19    | 971318  | (Z48053) putative<br>protein [Bovine<br>herpesvirus 1]   | 5.1     |
| 583       | D88000                                | Ralstonia<br>eutropha DNA<br>16S ribosomal<br>RNA > ::<br>dbj D88002 D88<br>002 Ralstonia<br>eutropha DNA<br>for 16S<br>ribosomal RNA<br>> ::<br>dbj D88003 D88<br>003 Ralstonia<br>eutropha DNA<br>for 16S<br>ribosomal RNA<br>> ::<br>dbj D88004 D88<br>004 Ralstonia<br>eutropha DNA<br>for 16S<br>ribosomal RNA | 0.19    | 3800952   | (AF100657) No<br>definition line found<br>[Caenorhabditis<br>elegans]  | 5.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 584    | U67462                                | Methanococcus jannaschii section 4 of 150 of the complete genome  | 0.19    | 3183617  | (AJ005586) MYB-related transcription factor [Antirrhinum majus]   | 4.0     |
| 585    | L23906                                | Gallus domesticus microsatellite DNA marker.  | 0.19    | 1947094  | (U93074) voltage-gated sodium channel homolog BdNa1   | 3.9     |
| 586    | AE001462                              | Helicobacter pylori, strain J99 section 23 of 132 of the complete genome  | 0.19    | 1730177  | GLUCOSE-6-PHOSPHATE ISOMERASE (GPI) ISOMERASE) (PHI) >gi 2118333 pir  I48073 glucose phosphate isomerase - Chinese hamster >gi 987046 griseus]  | 3.9     |
| 587    | M19460                                | P.putida catBC operon encoding cis,cis-muconate lactonizing enzyme I and muconolactone isomerase genes, complete cds. | 0.19    | 3873843  | (Z82256) cDNA EST yk251g7.3 comes from this gene; cDNA EST yk251g7.5 comes from this gene; cDNA EST EMBL:D68223 comes from this gene; cDNA EST EMBL:C12737 comes from this gene; cDNA EST yk389c8.5 comes from this gene; cDNA E... | 3.9     |
| 588    | U22349                                | Tetrahymena australis telomerase RNA gene, complete sequence  | 0.19    | 4105782  | (AF049922) PGP169-12 [Petunia x hybrida]  | 3.2     |
| 589    | L27745                                | Homo sapiens voltage operated calcium channel, alpha-1 subunit mRNA, complete cds.                                    | 0.19    | 3763926  | (AC004450) unknown protein [Arabidopsis thaliana]   | 3.0     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 590                                   | AF049588  | Canis familiaris<br>synapsin I gene,<br>partial cds  | 0.19    | 4104931   | (AF042196) auxin<br>response factor 8<br>[Arabidopsis thaliana]  | 3.0     |
| 591                                   | X06627    | Staphylococcus<br>aureus plasmid<br>pS194 sequence   | 0.19    | 137927  | PRE-NECK<br>APPENDAGE<br>PROTEIN (LATE<br>PROTEIN GP12)<br>>gi 75856 pir  WMBP<br>22 gene 12 protein -<br>phage phi-29<br>>gi 215330 (M14782)<br>pre-neck appendage<br>protein [Bacteriophage<br>phi-29]<br>>gi 225367 prf  13012<br>70G gene 12<br>[Bacteriophage phi-<br>29] | 2.3     |
| 592                                   | X61597    | M.musculus<br>gene for<br>kallikrein-<br>binding protein   | 0.19    | 2982874   | (AE000675)<br>cobalamin synthesis<br>related protein CobW  | 1.7     |
| 593                                   | AF016242  | Dictyostelium<br>discoideum<br>protein synthesis<br>elongation factor<br>1-alpha (tef2)<br>gene, partial cds   | 0.19    | 133659  | PUTATIVE RNA-<br>DIRECTED RNA<br>POLYMERASE  | 1.4     |
| 594                                   | AF004447  | Venezuelan<br>equine<br>encephalitis<br>virus strain 1327<br>polyprotein<br>gene, partial cds<br>> ::<br>gb AF004460 A<br>F004460<br>Venezuelan<br>equine<br>encephalitis<br>virus strain 1385<br>polyprotein<br>gene, partial cds | 0.19    | 4096173   | (U25968) early<br>embryogenesis protein<br>[Oryza sativa]  | 1.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 595    | J04821                                | Human elastin (ELN) gene, exon 1, clones HELC-5 and HELC-6.                 | 0.19    | 1170523  | INHIBIN BETA B CHAIN PRECURSOR inhibin precursor - bovine >gi 563753 (U16241) betaB inhibin/activin precursor [Bos taurus]  | 1.3     |
| 596    | AF059650                              | Homo sapiens histone deacetylase 3 (HDAC3) gene, complete cds               | 0.19    | 3024881  | PROBABLE TRANSPORT PROTEIN CY21C12.11 >gi 2078066 gnl PID e 315171 (Z95210) betP  | 0.83    |
| 597    | M69053                                | D.melanogaster calcium-activated K+ channel subunit                         | 0.19    | 1707984  | FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE I (FD-GOGAT) >gi 2126524 pir  S602 28 glutamate synthase (ferredoxin) (EC 1.4.7.1) gltB - Synechocystis sp. (PCC 6803) >gi 515938 (X80485) glutamate synthase | 0.80    |
| 598    | AF076279                              | Dictyostelium firmibasis plasmid Dfp1, complete plasmid sequence            | 0.19    | 453986   | (U00008) yejA [Escherichia coli]  | 0.79    |
| 599    | D28873                                | Mouse MCNP gene for C-type natriuretic peptide, complete cds (exon1, exon2) | 0.19    | 2650444  | (AE001092) acetyl-CoA synthetase (acs-1) [Archaeoglobus fulgidus]   | 0.63    |
| 600    | U06071                                | Oxytricha nova macronuclear actin II gene, complete cds.                    | 0.19    | 1584024  | complement control protein [Botryllus schlosseri]   | 0.48    |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|---|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 601                                   | L54057    | Homo sapiens CLP mRNA, partial cds.   | 0.19    | 3036883   | (AL022374) putative ABC transporter   | 0.46    |
| 602                                   | X89806    | P.lividius cDNA for COLL2alpha gene   | 0.19    | 3638957   | (AC004877) sco-spondin-mucin-like; similar to P98167 uncertain [Homo sapiens]                                     | 0.41    |
| 603                                   | AE001104  | Archaeoglobus fulgidus section 3 of 172 of the complete genome                            | 0.19    | 2315192   | (Y11739) transcription factor [Homo sapiens]  | 0.35    |
| 604                                   | U54501    | Rattus norvegicus microsatellite sequence D0Mco22   | 0.19    | 228951  | D-MeAsp receptor:ISOTYPE=epsilon3 [Mus musculus]  | 0.32    |
| 605                                   | X74468    | Human papillomavirus type 15 genomic DNA  | 0.19    | 3695390   | (AF096371) contains similarity to Rattus norvegicus cyclin G-associated kinase (SW:P97874) [Arabidopsis thaliana] | 0.28    |
| 606                                   | U20285    | Human Gps1 (GPS1) mRNA, complete cds  | 0.19    | 2582659   | (AJ002527) glucitol-6-phosphate dehydrogenase [Clostridium beijerinckii]  | 0.27    |
| 607                                   | D49408    | Human gene for interleukin 3 receptor alpha subunit, exon 10                              | 0.19    | 2522368   | (AF008596) alpha1,3-fucosyltransferase [Helicobacter pylori]  | 0.16    |
| 608                                   | AF041141  | Homo sapiens pituitary specific homeodomain protein (PROP1) gene, exon 3 and complete cds | 0.19    | 37403   | (X03541) trk gene product (aa 1-641) [Homo sapiens]   | 0.091   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 609    | L12531                                | Discopyge ommata Ca <sup>2+</sup> channel alpha 1 subunit gene sequence.                     | 0.19    | 3618274  | (AJ223219) hypothetical protein                           | 0.069   |
| 610    | AF052445                              | Yellow fever virus clone HONG9 polyprotein gene, complete cds                                | 0.19    | 1932822  | (U15928) KH-domain putative RNA binding protein           | 0.001   |
| 611    | Z36946                                | B.anthraxis sap gene encoding S-layer protein  | 0.19    | 173241   | (L06487) ZIP1 protein [Saccharomyces cerevisiae]          | 2e-04   |
| 612    | AF087984                              | Homo sapiens full length insert cDNA clone YW29A12   | 0.19    | 3786014  | (AC005499) hypothetical protein [Arabidopsis thaliana]    | 1e-06   |
| 613    | AE001010                              | Archaeoglobus fulgidus section 97 of 172 of the complete genome                              | 0.19    | 3135493  | (AF060248) unknown [Arabidopsis thaliana]                 | 7e-08   |
| 614    | L08965                                | Trichosporon cutaneum carbamoyl phosphate synthetase large subunit (argA) gene, partial cds. | 0.19    | 1086901  | (U41278) F33G12.3 gene product [Caenorhabditis elegans]   | 2e-08   |
| 615    | M91466                                | Rattus norvegicus A2b-adenosine receptor mRNA, complete cds.                                 | 0.19    | 2984320  | (AE000773) acetoin utilization protein [Aquifex aeolicus] | 6e-09   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 616       | X95971                                | S.lividans<br>groEL2 gene  | 0.19    | 3925277   | (AL032643) similar to<br>Uncharacterized<br>protein family<br>UPF0034, Double-<br>stranded RNA binding<br>motif; cDNA EST<br>yk489b3.5 comes from<br>this gene; cDNA EST<br>yk439g7.5 comes from<br>this gene<br>[Caenorhabditis<br>elegans] | 7e-10   |
| 617       | U12539                                | Schizosaccharo<br>myces pombe<br>scd2 (scd2)<br>gene, complete<br>cds.                                     | 0.19    | 1938549   | (U97016) similar to<br>drosophila Rlc1 gene<br>product ribosomal<br>protein L4 (YML4)<br>(NID:g459259)   | 3e-14   |
| 618       | U12539                                | Schizosaccharo<br>myces pombe<br>scd2 (scd2)<br>gene, complete<br>cds.                                     | 0.19    | 1938549   | (U97016) similar to<br>drosophila Rlc1 gene<br>product ribosomal<br>protein L4 (YML4)<br>(NID:g459259)   | 9e-15   |
| 619       | Z68327                                | Human DNA<br>sequence from<br>cosmid U25D11,<br>between markers<br>DXS366 and<br>DXS87 on<br>chromosome X. | 0.19    | 3875774   |  | 6e-15   |
| 620       | U66525                                | Dictyostelium<br>discoideum<br>ORFveg114<br>mRNA,<br>complete cds  | 0.19    | 3540281   | (AF056116) All-1<br>related protein [Fugu<br>rubripes]   | 2e-17   |
| 621       | U25830                                | Newcastle<br>disease virus<br>isolate Herts/33<br>matrix protein<br>mRNA,<br>complete cds                  | 0.19    | 2228750   | (U93868) RNA<br>polymerase III subunit<br>[Homo sapiens]   | 1e-18   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 622    | U89407                                | Mus musculus strain BALB/c delta-aminolevulinic acid dehydratase (Lv) mRNA, partial cds             | 0.19    | 1825764  | (U88314) C46H11.11 gene product [Caenorhabditis elegans] | 3e-25   |
| 623    | AF095598                              | Bison bison athabascaae microsatellite BBJ 2  | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 624    | AF064260                              | Strongylocentrotus purpuratus SRC8 mRNA, complete cds   | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 625    | U69533                                | Arabidopsis thaliana AtKAP alpha mRNA, complete cds   | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 626    | D89041                                | Bovine DNA for prostaglandin F2alpha receptor, partial cds  | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 627    | M24571                                | Dictyostelium discoideum tRNA-Glu-GAA gene, clone yGluGAA7.   | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 628    | X59772                                | D.melanogaster ovo gene required for female germ line development                                   | 0.18    | <NONE>   | <NONE>   | <NONE>  |
| 629    | AL010209                              | Plasmodium falciparum DNA ***<br>SEQUENCING IN PROGRESS<br>*** from contig 3-104, complete sequence | 0.18    | <NONE>   | <NONE>   | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 630    | U67575                                | Methanococcus jannaschii section 117 of 150 of the complete genome              | 0.18    | 111839   | inositol 1,4,5-triphosphate receptor 2 - rat  | 8.5     |
| 631    | U28730                                | Caenorhabditis elegans cosmid K10B2   | 0.18    | 1787604  | (AE000232) orf, hypothetical protein [Escherichia coli]   | 8.3     |
| 632    | X99798                                | L.lactis pepF1 & pepF2 genes  | 0.18    | 3406624  | (AF079110) glycosomal malate dehydrogenase [Trypanosoma brucei]   | 8.1     |
| 633    | AF025306                              | Danio rerio band 4.1-like protein 4 (nbl4) mRNA, complete cds                   | 0.18    | 465445   | PROBABLE NUCLEAR ANTIGEN herpesvirus 1 (strain Kaplan) >gi 334072 (M34651) ORF-3 protein [Pseudorabies virus] | 7.9     |
| 634    | AF059251                              | Mus musculus lipoxxygenase (alox) mRNA, complete cds                            | 0.18    | 1655667  | (Z81368) hypothetical protein Rv2393  | 6.6     |
| 635    | Z22605                                | G.domesticus CTCF protein mRNA.   | 0.18    | 481864   | 3-methyl-2-oxobutanoate dehydrogenase   | 6.6     |
| 636    | AB011086                              | Homo sapiens mRNA for KIAA0514 protein, complete cds                            | 0.18    | 3874158  | (Z81464) predicted using Genefinder   | 6.4     |
| 637    | Z78536                                | Caenorhabditis elegans cosmid C07A4, complete sequence [Caenorhabditis elegans] | 0.18    | 3702121  | (AJ011681) retinoblastoma-related protein [Chenopodium rubrum]  | 6.4     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 638    | U67530                                | Methanococcus jannaschii section 72 of 150 of the complete genome   | 0.18    | 3877946  | (Z81094) Weak similarity to 65 KDA heat shock protein (TR:G602231); cDNA EST EMBL:D71705 comes from this gene; cDNA EST EMBL:D74382 comes from this gene [Caenorhabditis elegans]   | 6.3     |
| 639    | M63781                                | Influenza A/Duck/England /1/62(H4N6) nucleoprotein mRNA, complete cds.  | 0.18    | 3873663  | (Z69634) cDNA EST EMBL:D71510 comes from this gene; cDNA EST EMBL:C08449 comes from this gene; cDNA EST yk266b12.3 comes from this gene; cDNA EST yk266b12.5 comes from this gene; cDNA EST yk461h7.3 comes from this gene; cDNA... | 6.2     |
| 640    | M73781                                | Oryctolagus cuniculus integrin beta-8 subunit mRNA, complete cds. >::gb I44828 I44828 Sequence 3 from patent US 5635601 | 0.18    | 1362129  | major allergen OLE17 - common olive   | 5.8     |
| 641    | X67219                                | D.melanogaster Rop gene   | 0.18    | 3449286  | (AB011527) MEGF1 [Rattus norvegicus]  | 4.8     |
| 642    | AF106941                              | Homo sapiens beta-arrestin 2 mRNA, complete cds   | 0.18    | 548353   | [PROTEIN-PII] URIDYLYLTRANSFERASE vinelandii >gi 39257 (X59610) uridylyl transferase  | 3.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 643    | AF052602                              | Danio rerio huntingtin (HD) mRNA, complete cds   | 0.18    | 241058   | potential IGF binding protein [chickens, Peptide Partial, 77 aa, segment 2 of 3]  | 3.6     |
| 644    | AB020709                              | Homo sapiens mRNA for KIAA0902 protein, complete cds                                     | 0.18    | 3875570  | (Z68314) predicted using Genefinder; cDNA EST EMBL:M75775 comes from this gene; cDNA EST EMBL:M89255 comes from this gene; cDNA EST EMBL:M89127 comes from this gene; cDNA EST EMBL:T00141 comes from this gene; cDNA EST EMBL:T... | 2.1     |
| 645    | AF096883                              | HIV-1 isolate patient 3 country USA pol polyprotein (pol) gene, partial cds              | 0.18    | 3250696  | (AL024486) putative protein   | 1.7     |
| 646    | L39928                                | Pyrocoelia miyako (clone pB-PmL41) luciferase mRNA, complete cds                         | 0.18    | 2914702  | (AC003974) unknown protein [Arabidopsis thaliana]   | 0.73    |
| 647    | M17082                                | Human carcinoembryonic nonspecific crossreacting antigen (CEA; NCA) gene, exons 1 and 2. | 0.18    | 1351833  | REGULATORY PROTEIN ABAA   | 0.72    |
| 648    | X75318                                | H.sapiens ITIH1 gene (exon 22) and ITIH3 gene  | 0.18    | 629557   | RNA-binding protein rnpD - Arabidopsis thaliana (fragment) >gi 510240 (X61108) RNA binding protein [Arabidopsis thaliana]   | 0.41    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 649    | AF011908                              | Mus musculus apoptosis associated tyrosine kinase (AATYK) mRNA, complete cds  | 0.18    | 330442   | (K03332) nuclear antigen 2 [Epstein-Barr virus]   | 5e-04   |
| 650    | U04004                                | Simian immunodeficiency virus SIVagmVER-2 envelope protein gene, partial cds. | 0.18    | 135102   | ASPARTYL-TRNA SYNTHETASE aspartate--tRNA ligase (EC 6.1.1.12) - Escherichia coli coli] >gi 1736513 gnl PID d1016401 (D90829) Aspartate--tRNA ligase (EC 6.1.1.12) [Escherichia coli]  | 6e-11   |
| 651    | U88155                                | Xenopus laevis RanGTPase activating protein                                   | 0.18    | 995714   | (X91258) pid:e198503 [Saccharomyces cerevisiae]   | 2e-13   |
| 652    | Z18921                                | B.oleracea gene for S-receptor kinase-like protein                            | 0.18    | 3875535  | (Z66511) similar to ribokinase; cDNA EST EMBL:D69553 comes from this gene; cDNA EST EMBL:D65938 comes from this gene; cDNA EST yk280h9.3 comes from this gene; cDNA EST yk280h9.5 comes from this gene; cDNA EST yk223d11.3 come... | 1e-19   |
| 653    | M60650                                | S.cerevisiae STA2 gene, complete cds.   | 0.16    | <NONE>   | <NONE>  | <NONE>  |
| 654    | U80912                                | Eucalyptus globulus NADP-isocitrate dehydrogenase (EgICDH) mRNA, complete cds | 0.16    | 3766172  | (AF057298) ornithine decarboxylase antizyme 2 [Mus musculus]  | 4.2     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 655                                   | AF012899  | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds  | 0.16    | 76749   | hypothetical protein 4 - fowl adenovirus 1  | 4.0     |
| 656                                   | AF027174  | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds                                   | 0.16    | 3044086   | (AF055904) unknown [Myxococcus xanthus]   | 0.60    |
| 657                                   | AF030231  | Glycine max sucrose synthase (SS) mRNA, complete cds   | 0.078   | <NONE>  | <NONE>  | <NONE>  |
| 658                                   | M19183    | Woodchuck hepatitis virus (WHV), complete genome, clone WHV 59.  | 0.072   | 1076190   | cell wall glycoprotein, 75K, precursor - diatom (Cylindrotheca fusiformis) >gi 515363 (X80394) P75K gene product [Cylindrotheca fusiformis] | 6.3     |
| 659                                   | U31557    | Ovine adenovirus IVa2 protein gene, DNA polymerase gene, terminal protein gene and 52.55 kDa protein gene, partial cds | 0.072   | 3511143   | (AF061244) unknown [Agrocybe aegerita]  | 6.2     |
| 660                                   | AL021491  | Caenorhabditis elegans cosmid Y44A6B, complete sequence [Caenorhabditis elegans]                                       | 0.070   | <NONE>  | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 661    | M33874                                | X.laevis Xotch protein mRNA, complete cds.   | 0.070   | 1654096  | (Y09076) RAD3 [Schizosaccharomyces pombe]   | 0.23    |
| 662    | AB012725                              | Mus musculus ZAN75 mRNA for zinc finger protein, complete cds  | 0.069   | 1350800  | MITOCHONDRIAL RIBOSOMAL PROTEIN S5  | 2.0     |
| 663    | AL021491                              | Caenorhabditis elegans cosmid Y44A6B, complete sequence [Caenorhabditis elegans]                     | 0.068   | <NONE>   | <NONE>  | <NONE>  |
| 664    | Z60318                                | H.sapiens CpG DNA, clone 1e1, reverse read cpg1e1.rla.   | 0.068   | 1280134  | (U55376) F16H11.2 gene product [Caenorhabditis elegans]   | 2.6     |
| 665    | Z35973                                | S.cerevisiae chromosome II reading frame ORF YBR104w   | 0.068   | 2493000  | PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR EMBL:Z14816 comes from this gene; cDNA EST EMBL:Z14946 comes from this gene; cDNA EST EMBL:D69746 comes from this gene; cDNA EST yk219b6.3 comes from this gene; cDNA ES... | 0.68    |
| 666    | Z86111                                | Streptomyces lividans rpsP, trmD, rplS, sipW, sipX, sipY, sipZ, mutT genes and 4 open reading frames | 0.068   | 1235974  | (X96713) collagen [Globodera pallida]   | 4e-04   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 667    | M72980                                | Anthonomus grandis vitellogenin gene (VTG), complete cds.   | 0.068   | 3242750  | (AC005164) match to ESTs AA731149 (NID:g2140138), AA731908 (NID:g2752719), AA287837 (NID:g1933519), AA262811 (NID:g1898382), and AA825820 (NID:g2899132) | 1e-59   |
| 668    | M34161                                | Rat tachykinin (PPT) gene, exons 5 and 6.   | 0.067   | <NONE>   | <NONE>   | <NONE>  |
| 669    | L03811                                | Aspergillus niger zinc finger protein (creA) gene, complete cds.                                      | 0.067   | <NONE>   | <NONE>   | <NONE>  |
| 670    | M64983                                | Human fibrinogen beta chain gene, complete mRNA. > gb I47706 I47706 Sequence 3 from patent US 5639940 | 0.067   | <NONE>   | <NONE>   | <NONE>  |
| 671    | AF014051                              | Nicotiana tabacum Mg chelatase subunit (ChlH) mRNA, partial cds                                       | 0.067   | <NONE>   | <NONE>   | <NONE>  |
| 672    | Y07540                                | H.sapiens sil gene  | 0.067   | 92331  | glycoprotein GP330, renal - rat (fragments)  | 7.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 673    | AJ000347                              | Rattus norvegicus mRNA for 3'(2'),5'-bisphosphate nucleotidase           | 0.067   | 129238   | 25 KD OOKINETE SURFACE ANTIGEN PRECURSOR (PRS25)<br>>gi 320962 pir  A4496625k ookinete surface antigen precursor - Plasmodium reichenowi reichenowi]                               | 7.4     |
| 674    | L19979                                | Squid sodium channel mRNA, complete cds.                                 | 0.067   | 2128473  | hypothetical protein MJ0750 - Methanococcus jannaschii<br>>gi 1592304 (U67521) ferredoxin-type protein   | 1.5     |
| 675    | X08050                                | Yeast tRNA-Glu(3) gene and flanking regions                              | 0.067   | 1334398  | (X15081) MURF2 protein (AA 1-348)  | 0.65    |
| 676    | X17115                                | Human mRNA for IgM heavy chain complete sequence                         | 0.067   | 1731331  | HYPOTHETICAL 51.6 KD PROTEIN CY49.14C<br>>gi 1370241 gnl PID e247089 (Z73966) hypothetical protein Rv2075c [Mycobacterium tuberculosis]  | 0.51    |
| 677    | AF032871                              | Homo sapiens uncoupling protein 3 (UCP3) gene, exon 1 and partial exon 2 | 0.067   | 112900   | ALPHA-2C-1 ADRENERGIC RECEPTOR human<br>>gi 178194 (J03853) kidney alpha-2-adrenergic receptor [Homo sapiens]<br>>gi 1628638 (U72648) alpha2-C4-adrenergic receptor [Homo sapiens] | 0.50    |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 678    | X05319                                | Mouse class II MHC E-beta 2 (d) gene exon 3                                       | 0.067   | 585074   | DYNAMIN 3 (DYNAMIN, TESTICULAR) rat >gi 391872 gnl PID d1003668 (D14076) testicular dynamin [Rattus norvegicus] | 3e-04   |
| 679    | AB006362                              | Candida albicans CaSLN1 gene, complete cds  | 0.067   | 3417296  | (AC003007) Unknown gene product (partial) [Homo sapiens]  | 9e-56   |
| 680    | AF021236                              | African horse sickness virus capsid VP3 (L3) mRNA, complete cds                   | 0.066   | <NONE>   | <NONE>  | <NONE>  |
| 681    | AE001507                              | Helicobacter pylori, strain J99 section 68 of 132 of the complete genome          | 0.066   | <NONE>   | <NONE>  | <NONE>  |
| 682    | AF039717                              | Caenorhabditis elegans cosmid R13H8   | 0.066   | <NONE>   | <NONE>  | <NONE>  |
| 683    | AF029027                              | Syncerus caffer isolate Queen Elizabeth Mweya 14 mitochondrial DNA control region | 0.066   | <NONE>   | <NONE>  | <NONE>  |
| 684    | AF087967                              | Homo sapiens full length insert cDNA clone YU51G05                                | 0.066   | 2982476  | (X97203) C1 protein [Beet curly top virus]  | 9.5     |
| 685    | J02037                                | Baboon endogenous virus proviral long terminal repeat DNA.                        | 0.066   | 972767   | (L37868) POU-domain transcription factor [Homo sapiens]   | 7.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 686    | AF000141                              | Lycopersicon esculentum class I knotted-like homeodomain protein (LeT6) mRNA, complete cds                    | 0.066   | 3157926  | (AC002131) Strong similarity to extensin-like protein gb Z34465 from Zea mays. [Arabidopsis thaliana] | 5.6     |
| 687    | AB001746                              | Bensingtonia sp. OK255 gene for 18S rRNA > :: dbj AB001747 A B001747 Bensingtonia sp. OK259 gene for 18S rRNA | 0.066   | 3859889  | (AF070064) cap 'n' collar isoform C [Drosophila melanogaster]   | 0.38    |
| 688    | AE001461                              | Helicobacter pylori, strain J99 section 22 of 132 of the complete genome                                      | 0.065   | <NONE>   | <NONE>  | <NONE>  |
| 689    | M30821                                | Chicken erythroid transport proteins c1 and c2  | 0.065   | <NONE>   | <NONE>  | <NONE>  |
| 690    | AB009802                              | Homo sapiens gene for osteonidogen, intron 3  | 0.065   | <NONE>   | <NONE>  | <NONE>  |
| 691    | AF086062                              | Homo sapiens full length insert cDNA clone YZ06B11  | 0.065   | <NONE>   | <NONE>  | <NONE>  |
| 692    | AB002369                              | Human mRNA for KIAA0371 gene, complete cds  | 0.065   | 2500884  | SIGNAL SEQUENCE BINDING PROTEIN binding protein [Synechococcus sp.]                                   | 5.5     |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 693                                   | AF086864  | Cyclopodia sp.<br>large subunit<br>ribosomal RNA<br>gene,<br>mitochondrial<br>gene for<br>mitochondrial<br>RNAs, partial<br>sequence > ::<br>gb AF086866 A<br>F086866<br>Penicillidia sp.<br>large subunit<br>ribosomal RNA<br>gene,<br>mitochondrial<br>gene for<br>mitochondrial<br>RNAs, partial<br>sequence | 0.065   | 3721684   | (AB012957) probable<br>glycosyl transferase<br>[Vibrio cholerae] | 5.5     |
| 694                                   | L44593    | Bacteriophage<br>BK5-TORF'410,<br>3' end pf cds, 20<br>ORFs, repressor<br>protein, and Cro<br>repressor protein<br>genes, complete<br>cds, ORF70'<br>gene, 5' end of<br>cds.  | 0.065   | 1172067   | PEPTIDASE T<br>(AMINOTRIPEPTID<br>ASE) influenzae Rd]            | 3.2     |
| 695                                   | U80079    | Ciona<br>intestinalis<br>MyoD-family<br>protein<br>(CiMDFa)<br>mRNA,<br>complete cds  | 0.065   | 4218110   | (AL035353) contains<br>EST gb:F15281                             | 2.5     |
| 696                                   | AB020718  | Homo sapiens<br>mRNA for<br>KIAA0911<br>protein,<br>complete cds  | 0.065   | 1722734   | MINOR CAPSID<br>PROTEIN L2<br>>gi 1020192 type 23]               | 1.9     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 697    | AF082137                              | Zea mays copia-like retrotransposon Stl-14 leader region, partial sequence                                 | 0.065   | 1877501  | (U89278) polyhomeotic 2 homolog [Homo sapiens]                     | 1.1     |
| 698    | X64053                                | R.norvegicus ZnBP gene for zinc binding protein  | 0.065   | 464963   | TRYPSIN PRECURSOR  | 0.36    |
| 699    | U67065                                | Mus musculus butyrophilin (BTN) gene, promoter region and complete cds                                     | 0.065   | 2132252  | hypothetical protein YPL263c - yeast                               | 3e-10   |
| 700    | M64862                                | Rat matrin F/G mRNA, complete cds.   | 0.065   | 3420183  | (AF041105) organic anion transporter protein 3 [Rattus norvegicus] | 4e-19   |
| 701    | K02205                                | Yeast (S.cerevisiae) transcriptional activator of amino acid-biosynthetic genes (GCN4) gene, complete cds. | 0.064   | <NONE>   | <NONE>   | <NONE>  |
| 702    | X58282                                | Maize mRNA for a high mobility group protein   | 0.064   | <NONE>   | <NONE>   | <NONE>  |
| 703    | AC001545                              | Homo sapiens (subclone 1_f3 from P1 H69) DNA sequence  | 0.064   | <NONE>   | <NONE>   | <NONE>  |
| 704    | AF023461                              | Homo sapiens FRA3B region sequence   | 0.064   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 705    | U50307                                | Caenorhabditis elegans cosmid F43H9.   | 0.064   | <NONE>   | <NONE>  | <NONE>  |
| 706    | U46542                                | Streptococcus crista HmpA gene, partial cds, putative adhesin/ABC transport system protein (scbA) gene, complete cds | 0.064   | 1209391  | (D83659) TPR protein pombe]<br>>gi 2894282 gnl PID e1251103 (AL021838) pre-mrna splicing factor. [Schizosaccharomyces pombe]  | 9.2     |
| 707    | X57564                                | A.rusticana mRNA for neutral peroxidase  | 0.064   | 1492037  | (U60315) MC094R [Molluscum contagiosum virus subtype 1]   | 6.9     |
| 708    | U06986                                | Human alpha-2-macroglobulin receptor/lipoprotein receptor protein (A2MR/LRP) gene, exons 39-41.                      | 0.064   | 100800   | rab15B protein - wheat >gi 21853 (X62476) rab protein [Triticum aestivum]   | 5.3     |
| 709    | D85773                                | Human CpG island sequence, clone Q28B8   | 0.064   | 2245382  | (U88325) suppressor of cytokine signalling-1 [Mus musculus]   | 5.3     |
| 710    | L06178                                | Apis mellifera ligustica complete mitochondrial genome   | 0.064   | 3695379  | (AF096370) contains similarity to a C. elegans hypothetical protein F44G4.1 (GB:Z49910) and several yeast hypothetical proteins such as 35.1 KD protein in NAM8-GAR1 intergenic region (SP:P38805) [Arabidopsis thaliana] | 3.2     |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|---|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 711       | Y16242                                | Triticum aestivum mRNA for beta-amylase   | 0.064   | 1175958   | HYPOTHETICAL<br>70.5 KD PROTEIN IN<br>AGP3-DAK3<br>INTERGENIC<br>REGION<br>>gi 1084712 pir  S562<br>01 probable membrane<br>protein YFL054c -<br>yeast (Saccharomyces<br>cerevisiae)<br>>gi 836701 gnl PID d1<br>009825 (D50617)<br>YFL054C | 3.1     |
| 712       | L81779                                | Homo sapiens<br>(subclone 2_a2<br>from P1 H25)<br>DNA sequence  | 0.064   | 3845169   | (AE001391)<br>phosphatase (acid<br>phosphatase family)  | 0.81    |
| 713       | X13826                                | C.reinhardtii<br>psb1 mRNA for<br>OEE1 protein of<br>photosystem II<br>(oxygen-<br>evolving<br>enhancer<br>protein) | 0.064   | 171040  | (M94535) ATPase<br>[Saccharomyces<br>cerevisiae] cerevisiae,<br>Peptide, 377 aa]<br>[Saccharomyces<br>cerevisiae]   | 0.054   |
| 714       | X06487                                | H.sapiens<br>mRNA for bc12-<br>Ig fusion gene   | 0.064   | 2429362   | (AF020261) proline<br>rich protein [Santalum<br>album]  | 0.016   |
| 715       | U79638                                | Mus musculus<br>cyclin-dependent<br>kinase inhibitor<br>protein<br>(p15(INK4b))<br>gene, exon 2 and<br>partial cds  | 0.064   | 3929221   | (AF082557) TRF1-<br>interacting ankyrin-<br>related ADP-ribose<br>polymerase [Homo<br>sapiens]  | 1e-10   |
| 716       | U39099                                | Human T cell<br>receptor alpha<br>chain mRNA,<br>partial cds  | 0.063   | <NONE>  | <NONE>  | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 717                                   | U39673    | Clostridium acetobutylicum KdpC (kdpC) gene, partial cds, sensor histidine kinase homolog (kdpD) and response regulator homolog (kdpE) genes, complete cds                                     | 0.063   | <NONE>  | <NONE>  | <NONE>  |
| 718                                   | AL022317  | Human DNA sequence from clone 140L1 on chromosome 22q13.1-13.31, complete sequence [Homo sapiens]  | 0.063   | 1931640   | (U95973) Serine carboxypeptidase isolog [Arabidopsis thaliana]                | 5.2     |
| 719                                   | U28972    | Spiroplasma citri orfa and orff genes, partial cds, orfb, orfc, and orfe genes and Spiroplasma virus SpV1-derived ORF1 and ORF3 genes, complete cds, and SpV1-derived ORF14 gene, partial cds. | 0.063   | 4091939   | (AF070704) envelope glycoprotein [Human immunodeficiency virus type 1]        | 5.2     |
| 720                                   | U15159    | Mus musculus limk kinase (limk) mRNA, complete cds   | 0.063   | 3638957   | (AC004877) sco-spondin-mucin-like; similar to P98167 uncertain [Homo sapiens] | 5.1     |
| 721                                   | AF058416  | Homo sapiens lipoprotein receptor-related protein (LRP1), exons 39, 40, and 41   | 0.063   | 1788123   | (AE000276) orf, hypothetical protein [Escherichia coli]                       | 4.0     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 722                                   | AE001430  | Plasmodium falciparum chromosome 2, section 67 of 73 of the complete sequence  | 0.063   | 2244849   | (Z97337) hypothetical protein   | 4.0     |
| 723                                   | L29323    | Streptococcus pneumoniae methyl transferase gene cluster, complete sequence  | 0.063   | 3874022   | (Z70203) cDNA EST EMBL:D72339 comes from this gene; cDNA EST EMBL:D75197 comes from this gene [Caenorhabditis elegans]  | 2.3     |
| 724                                   | X72631    | H.sapiens mRNA encoding Rev-ErbAalpha > :: emb X72632 HS REVERB2 H.sapiens mRNA encoding Rev-ErbAalpha (internal fragment) | 0.063   | 3979878   | (Z73105) predicted using Genefinder; cDNA EST EMBL:T01277 comes from this gene; cDNA EST EMBL:T01796 comes from this gene; cDNA EST EMBL:D32545 comes from this gene; cDNA EST EMBL:D33060 comes from this gene; cDNA EST EMBL:D... | 1.7     |
| 725                                   | U17969    | Human initiation factor eIF-5A gene, complete cds.   | 0.063   | 2429509   | (AF025467) contains similarity to drosophila DNA-binding protein K10 (NID:g8148) [Caenorhabditis elegans]   | 1.4     |
| 726                                   | AE001000  | Archaeoglobus fulgidus section 107 of 172 of the complete genome   | 0.063   | 3462802   | (AF082486) nef protein [Human immunodeficiency virus type 1]  | 0.35    |



| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|---|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 727                                   | S80986    | svp[40]=svp-related nuclear receptor/retinoid signaling modulator [zebrafishes, mRNA, 3876 nt]  | 0.063   | 1326288   | (U58734) weak similarity to ankyrin G [Caenorhabditis elegans]  | 0.093   |
| 728                                   | AF109134  | Homo sapiens 7-60 mRNA, complete cds  | 0.063   | 1083764   | proline-rich proteoglycan 2 precursor, parotid - rat >gi 310200 (L17318) proline-rich proteoglycan [Rattus norvegicus]            | 0.001   |
| 729                                   | D87466    | Human mRNA for KIAA0276 gene, partial cds   | 0.063   | 2879865   | (AL021816) SPBC24E9.03c, unknown, len:251aa [Schizosaccharomyces pombe]   | 6e-05   |
| 730                                   | AB018269  | Homo sapiens mRNA for KIAA0726 protein, complete cds  | 0.063   | 2995865   | (AF053455) tetraspan TM4SF [Homo sapiens]   | 2e-16   |
| 731                                   | D86954    | Cricetulus griseus mRNA for Cytochrome P-450 2A14, complete cds                                 | 0.063   | 2496896   | HYPOTHETICAL 47.6 KD PROTEIN C16C10.5 IN CHROMOSOME III >gi 3874383 gnl PID e 1344077 type (RING finger) [Caenorhabditis elegans] | 1e-22   |
| 732                                   | AL010232  | Plasmodium falciparum DNA ***<br>SEQUENCING IN PROGRESS *** from contig 4-58, complete sequence | 0.062   | <NONE>  | <NONE>  | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 733                                   | U90714    | Mycoplasma gallisepticum haemagglutinin precursor genes, complete cds  | 0.062   | <NONE>  | <NONE>  | <NONE>  |
| 734                                   | AF107044  | Homo sapiens clone pCL4 DNA-binding protein SOX21 (SOX21) gene, complete cds                                 | 0.062   | <NONE>  | <NONE>  | <NONE>  |
| 735                                   | L41729    | Caenorhabditis elegans Ro ribonucleoprotein autoantigen mRNA, complete cds                                   | 0.062   | 2983060   | (AE000687) putative protein [Aquifex aeolicus]  | 8.6     |
| 736                                   | Z99287    | Caenorhabditis elegans cosmid Y7A9D, complete sequence [Caenorhabditis elegans]                              | 0.062   | 1176542   | PUTATIVE SERINE/THREONINE-PROTEIN KINASE D1044.3 IN CHROMOSOME III >gi 495684 (U00065) contains EGF-like repeats; highly similar to ZC84.1; 3' exons similar to protein kinase [Caenorhabditis elegans] | 5.8     |
| 737                                   | AB014514  | Homo sapiens mRNA for KIAA0614 protein, partial cds  | 0.062   | 4033395   | DNA GYRASE SUBUNIT B subunit [Myxococcus xanthus]   | 3.9     |
| 738                                   | L29165    | Human germline immunoglobulin light chain variable region (lambda-IIIb subgroup) from IgM rheumatoid factor. | 0.062   | 1914685   | (Y12014) RAD23 protein, isoform II  | 1.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 739    | U09364                                | Schistosoma japonicum Chinese clone pY6 paramyosin mRNA, partial cds.            | 0.062   | 1350800  | MITOCHONDRIAL RIBOSOMAL PROTEIN S5   | 1.3     |
| 740    | Y16242                                | Triticum aestivum mRNA for beta-amylase  | 0.062   | 79834  | hypothetical protein 1246 (uvrA region) - Micrococcus luteus (fragment)  | 0.59    |
| 741    | M97695                                | Leishmania pifanoi cysteine proteinase (cys2) gene, complete cds.                | 0.062   | 1174754  | TROPOMYOSIN 1 (TMI) (POLYPEPTIDE 49) >gi 320989 pir  A60607 tropomyosin - fluke  | 0.018   |
| 742    | U67526                                | Methanococcus jannaschii section 68 of 150 of the complete genome                | 0.062   | 1330345  | (U58755) coded for by C. elegans cDNA yk34b1.5; coded for by C. elegans cDNA yk13h10.5; coded for by C. elegans cDNA yk46e8.5; coded for by C. elegans cDNA yk46d5.5; coded for by C. elegans cDNA yk43c2.5; coded for by C. elegans cDNA yk46e8.... | 1e-40   |
| 743    | Z78414                                | Caenorhabditis elegans cosmid W09D12, complete sequence [Caenorhabditis elegans] | 0.061   | <NONE>   | <NONE>   | <NONE>  |
| 744    | Y13606                                | Mus musculus gene encoding filensin, exons 6, 7                                  | 0.061   | 2314715  | (AE000651) H. pylori predicted coding region HP1527  | 4.9     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 745    | J04374                                | Eggplant mosaic virus genome.   | 0.061   | 141449   | HYPOTHETICAL 35.5 KD PROTEIN IN TRANSPOSON TN4556<br>>gi 80759 pir  JQ0431<br>hypothetical 35.5K protein - Streptomyces fradiae transposon Tn4556 | 3.8     |
| 746    | AB022200                              | Marine obligately oligotrophic bacterium POO-10 DNA for 16S ribosomal RNA, partial sequence | 0.061   | 3983593  | (AB000307) transcarboxylase-beta  | 2.2     |
| 747    | X54250                                | Rat mRNA for zinc finger protein AT-BP2, partial cds  | 0.061   | 1377886  | (L46815) DNA binding protein Rc [Mus musculus]  | 0.98    |
| 748    | X69942                                | M.musculus mRNA of enhancer-trap-locus 1  | 0.061   | 2983969  | (AE000748) putative protein [Aquifex aeolicus]  | 0.57    |
| 749    | AJ223206                              | Mus musculus mRNA for scrapie responsive protein 1  | 0.061   | 4204265  | (AC005223) 45643 [Arabidopsis thaliana]   | 5e-31   |
| 750    | Y10205                                | H.sapiens mRNA for CD88 protein   | 0.060   | <NONE>   | <NONE>  | <NONE>  |
| 751    | U79260                                | Human clone 23745 mRNA, complete cds  | 0.060   | <NONE>   | <NONE>  | <NONE>  |
| 752    | X07453                                | Plasmodium falciparum 11-1 gene part 1  | 0.060   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION                                    | P VALUE |
| 753    | U57502                                | Rattus norvegicus protein tyrosine phosphatase delta gene, catalytic domain, partial cds. | 0.060   | 3452285  | (AF044915) polar tube protein PTP55 precursor  | 0.28    |
| 754    | X68359                                | M.fascicularis gene for apolipoprotein C-III  | 0.060   | 730843   | SHUTTLE CRAFT PROTEIN >gi 487400               | 2e-04   |
| 755    | X51634                                | Pseudomonas braB gene for branched chain amino acid transport carrier (LIV-II)            | 0.059   | 1835622  | (U85718) CCML [Pseudomonas putida GB-1]        | 8.1     |
| 756    | AF072405                              | Gossypium hirsutum cotton fiber expressed protein 2 (CFE2) mRNA, complete cds             | 0.059   | 423766   | alkaline phosphatase, 145K - Synechococcus sp. | 4.7     |
| 757    | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                 | 0.056   | 2662481  | (AF034859) juvenile hormone resistance protein | 3.3     |
| 758    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds   | 0.054   | 547847   | LECTIN PRECURSOR                               | 7.0     |
| 759    | X61046                                | Hydra N-COL 2 mRNA for mini-collagen, partial cds   | 0.053   | <NONE>   | <NONE>   | <NONE>  |
| 760    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                                  | 0.052   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 761    | S79843                                | {random amplified hybridization microsatellite RAHM} [Beta vulgaris=sugar beets, Genomic, 537 nt] | 0.025   | 1730145  | GAMETOGENESIS EXPRESSED PROTEIN GEG-154 >gi 2137331 pir  I4836 1 gene GEG-154 protein - mouse >gi 550123 (X71642) pid:g550123 [Mus musculus] | 2e-16   |
| 762    | AB000096                              | Mouse mRNA for GATA-2 protein, complete cds   | 0.023   | <NONE>   | <NONE>   | <NONE>  |
| 763    | Z62366                                | H.sapiens CpG DNA, clone 67h7, forward read cpg67h7.fila.   | 0.023   | 3123312  | ZINC FINGER PROTEIN 142 (KIAA0236) to Human zinc finger protein(ZNF142) [Homo sapiens]   | 5.9     |
| 764    | L11670                                | Human transmembrane glycoprotein (CD53) gene, exons 2 through 8.                                  | 0.023   | 80636  | hypothetical 67K protein - Mycobacterium fortuitum plasmid pAL5000 >gi 149986 (M60875) ORF2  | 3.4     |
| 765    | D83984                                | Sulculus diversicolor DNA for IDO-like myoglobin, complete cds                                    | 0.023   | 3114665  | (AF061267) inner membrane component HtxE [Pseudomonas stutzeri]  | 3.4     |
| 766    | X98890                                | S.tuberosum mRNA for inorganic phosphate transporter, StPT1                                       | 0.023   | 683532   | (X02155) thyroglobulin [Bos taurus]  | 1.1     |
| 767    | U58835                                | Dissostichus mawsoni preprotrypsin gene, complete cds   | 0.022   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|---|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 768    | AJ009630                              | Glomus versiforme chitin synthase gene (clone Gvchs3)   | 0.022   | <NONE>   | <NONE>      | <NONE>  |
| 769    | J04040                                | Human glucagon mRNA, complete cds.  | 0.022   | <NONE>   | <NONE>      | <NONE>  |
| 770    | X74908                                | L.esculentum Asr3 gene  | 0.022   | <NONE>   | <NONE>      | <NONE>  |
| 771    | L07293                                | Shigella dysenteriae O-antigen polysaccharide biosynthesis rfbX, O-antigen polymerase (rfc), rhamnosyl transferase I and II (rfbR and rfbQ) and rfbD genes, complete cds. | 0.022   | <NONE>   | <NONE>      | <NONE>  |
| 772    | AF040094                              | Mus musculus inositol polyphosphate 5-phosphatase II (INPP5P) mRNA, complete cds  | 0.022   | <NONE>   | <NONE>      | <NONE>  |
| 773    | X76776                                | H.sapiens HLA-DMB gene  | 0.022   | <NONE>   | <NONE>      | <NONE>  |
| 774    | AE001521                              | Helicobacter pylori, strain J99 section 82 of 132 of the complete genome  | 0.022   | <NONE>   | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 775    | X16004                                | A.longa rbcL, rpl5, rps8, rpl36, rps14, rps2, trnI, trnF, trnC and rpoB (partial) genes > :: emb X75651 AL RIBPA.longa plastid genes for ribosomal proteins, tRNAs, RNA polymerase subunit beta and rubisco large subunit | 0.022   | <NONE>   | <NONE>  | <NONE>  |
| 776    | Y12707                                | Lactococcus lactis cremoris plasmid pHW393 DNA, rlladii, mlladii genes  | 0.022   | <NONE>   | <NONE>  | <NONE>  |
| 777    | U27118                                | Arabidopsis thaliana glutamyl-tRNA reductase  | 0.022   | <NONE>   | <NONE>  | <NONE>  |
| 778    | Z96622                                | H.sapiens telomeric DNA sequence, clone 5PTEL002, read 5PTELOO002.seq   | 0.022   | 191333   | (J05503) carbamoyl-phosphate synthetase (E.C.6.3.5.5)         | 9.8     |
| 779    | D83984                                | Sulculus diversicolor DNA for IDO-like myoglobin, complete cds  | 0.022   | 1078509  | probable membrane protein YDR018c - yeast                     | 9.7     |
| 780    | Z77952                                | H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA4A3   | 0.022   | 4204206  | (AB022786) N-acetyl-beta-D-glucosaminidase [Enterobacter sp.] | 7.5     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 781    | M10217                                | Xenopus laevis mitochondrial DNA, complete genome.                                  | 0.022   | 2145763  | B2168_C2_205 protein - Mycobacterium leprae  | 7.3     |
| 782    | M55147                                | Pea chloroplast glyceraldehyde-3-phosphate dehydrogenase (Gpb1) gene, complete cds. | 0.022   | 417308   | PROBABLE HELICASE MOT1 Mot1p is a probable helicase essential for vegetative growth on rich glucose medium at 30 degree C: Swiss-Prot Accession number P32333; similar to S. cerevisiae RAD26 gene product: Swiss-Prot Accession number P40352 | 4.2     |
| 783    | X58839                                | Acholeplasma virus MV-L1 DNA for complete circular genome                           | 0.022   | 3273189  | (AB008757) subunit II of c(o/b)3-type cytochrome c oxidase [Bacillus stearothermophilus]   | 4.1     |
| 784    | M26185                                | Mouse c-myb oncogene, exon 1 and exon 2 (partial).                                  | 0.022   | 138592   | VITELLOGENIN I PRECURSOR (YOLK PROTEIN 1) >gi 72270 pir  VJFF1 vitellogenin I precursor unnamed protein product [Drosophila melanogaster]  | 2.5     |
| 785    | AF061195                              | Streptomyces albus valine dehydrogenase (Vdh) gene, complete cds                    | 0.022   | 2088768  | (AF003145) B0414.8 gene product [Caenorhabditis elegans]   | 0.86    |
| 786    | AF053622                              | Homo sapiens alpha 1,2-mannosidase IB gene, exon 9                                  | 0.022   | 1352361  | EARLY GROWTH RESPONSE PROTEIN 1 fish >gi 531456 (U12895) egr1 [Danio rerio]  | 0.36    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 787    | Z71500                                | S.cerevisiae chromosome XIV reading frame ORF YNL224c  | 0.022   | 1708875  | PUTATIVE TUMOR SUPPRESSOR LUCA15 sapiens]                            | 0.16    |
| 788    | D10471                                | Herpes simplex virus type 2 genomic DNA for 0.74-0.84 region, complete cds                                 | 0.022   | 3132276  | (AB011486) short ORF [TT virus]                                      | 0.13    |
| 789    | U43082                                | Zea mays T cytoplasm male sterility restorer factor 2 (rf2) mRNA, complete cds                             | 0.022   | 3319720  | (AL031035) putative aldehyde dehydrogenase [Streptomyces coelicolor] | 0.011   |
| 790    | X86913                                | H.sapiens simple tandem repeat DNA (clone wg3a6)   | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 791    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 792    | U34016                                | Nannostomus sp. large subunit rRNA gene, mitochondrial gene encoding mitochondrial rRNA, partial sequence. | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 793    | X00845                                | Yeast mitochondrial genes for 15S rRNA and tRNA-Trp  | 0.021   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 794    | AB012113                              | Homo sapiens gene for CC chemokine PARC precursor, complete cds  | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 795    | U62395                                | Daucus carota globulin-like protein (Gea8) gene, complete cds  | 0.021   | <NONE>   | <NONE>   | <NONE>  |
| 796    | M22718                                | P.falciparum actin II gene, complete cds.  | 0.021   | 2623773  | (AF004835) tyrocidine synthetase 3 [Brevibacillus brevis]  | 8.8     |
| 797    | U27118                                | Arabidopsis thaliana glutamyl-tRNA reductase   | 0.021   | 3549885  | (AJ006631) cysteine-rich secretory protein-1 [Equus caballus]  | 8.8     |
| 798    | X99832                                | H.sapiens CLN3 gene, complete CDS  | 0.021   | 262249   | (S52010) orf1 5' of EpoR [mice, Peptide, 85 aa] [Mus sp.]  | 8.7     |
| 799    | AF016266                              | Homo sapiens TRAIL receptor 2 mRNA, complete cds   | 0.021   | 729048   | SUCCINYL-COA:COENZYME A TRANSFERASE transferase [Clostridium kluyveri]   | 8.7     |
| 800    | Z92541                                | Human DNA sequence from PAC 179I15, BRCA2 gene region chromosome 13q12-13 contains lactase-phlorizin hydrolase (LCT) | 0.021   | 585820   | LIPOPOLYSACCHARIDE 1,2-N-ACETYLGLUCOSAMINETRANSFERASE >gi 466761 (U00039) rfaK [Escherichia coli] >gi 1790053 (AE000440) probably hexose transferase; lipopolysaccharide core biosynthesis | 5.3     |
| 801    | S58588                                | dopamine D2 receptor [human, brain, Genomic, 3794 nt, segment 4 of 5]  | 0.021   | 2677620  | (Y08029) NAD(P)(+)-arginine ADP-ribosyltransferase [Oryctolagus cuniculus]   | 5.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 802    | M60522                                | Rat nerve growth factor-inducible protein (VGF) gene, complete cds. | 0.021   | 4103934  | (AF030050) replication factor C [Rattus norvegicus]  | 3.1     |
| 803    | AF045654                              | Gallus gallus neuregulin beta-la mRNA, complete cds                 | 0.021   | 2746829  | (AF040647) No definition line found [Caenorhabditis elegans]                                 | 3.0     |
| 804    | M69023                                | Human globin gene.  | 0.021   | 3880259  | (Z82056) T26H5.8 [Caenorhabditis elegans]<br>>gi 3880787 gnl PID e1350288 (AL032620) T26H5.8 | 2.4     |
| 805    | Z65960                                | H.sapiens CpG DNA, clone 69d2, reverse read cpg69d2.rtlb.           | 0.021   | 1707245  | (U80845) similar to family 1 of G-protein coupled receptors [Caenorhabditis elegans]         | 0.79    |
| 806    | X97073                                | A.oligospora gene encoding lectin                                   | 0.021   | 116949   | CORE ANTIGEN >gi 73601 pir  NKVLC2 core antigen - woodchuck hepatitis virus 2 >gi 336135     | 0.47    |
| 807    | X56491                                | D. melanogaster mRNA for gene containing opa repetitive element     | 0.021   | 2842750  | HOMEBOX PROTEIN DLX-7 >gi 1620520  | 0.16    |
| 808    | L78760                                | Homo sapiens (subclone 1_f6 from P1 H31) DNA sequence               | 0.021   | 113671   | !!!! ALU CLASS F WARNING ENTRY !!!!  | 0.15    |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 809       | AB007864                              | Homo sapiens<br>KIAA0404<br>mRNA, partial<br>cds   | 0.021   | 118144  | CYSTEINE<br>SYNTHASE A (O-<br>ACETYL SERINE<br>SULFHYDRYLASE<br>A) (O-<br>ACETYL SERINE<br>(THIOL)- LYASE A)<br>(CSASE A)<br>>gi 68323 pir  SYEBA<br>C cysteine synthase<br>(EC 4.2.99.8) A -<br>Salmonella<br>typhimurium<br>>gi 153935 (M21450)<br>cysK protein<br>[Salmonella<br>typhimurium] | 0.12    |
| 810       | AL021932                              | Mycobacterium<br>tuberculosis<br>H37Rv complete<br>genome;<br>segment 22/162                               | 0.021   | 2909514   | (AL021932)<br>hypothetical protein<br>Rv0439c  | 7e-10   |
| 811       | U89991                                | Hypocrea<br>jecorina<br>mannose-1-<br>phosphate<br>guanylyltransfer<br>ase (MPG1)<br>mRNA,<br>complete cds | 0.021   | 3581924   | (AL031538) mannose-<br>1-phosphate<br>guanylyltransferase<br>[Schizosaccharomyces<br>pombe]  | 6e-20   |
| 812       | X00641                                | Sugar beet<br>mitochondrial<br>minicircle pO<br>sequence   | 0.020   | <NONE>  | <NONE>   | <NONE>  |
| 813       | Z50097                                | D.melanogaster<br>mRNA for hdc<br>protein.   | 0.020   | <NONE>  | <NONE>   | <NONE>  |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|---|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 814       | AF044866                              | Phoebis sennae<br>large subunit<br>ribosomal RNA<br>gene, partial<br>sequence; tRNA-<br>Val gene,<br>complete<br>sequence; and<br>small subunit<br>ribosomal RNA<br>gene, partial<br>sequence,<br>mitochondrial<br>genes for<br>mitochondrial<br>RNAs | 0.020   | <NONE>  | <NONE>   | <NONE>  |
| 815       | AF074386                              | Sambucus nigra<br>hevein-like<br>protein mRNA,<br>complete cds  | 0.020   | <NONE>  | <NONE>   | <NONE>  |
| 816       | AF027174                              | Arabidopsis<br>thaliana<br>cellulose<br>synthase<br>catalytic subunit<br>(Ath-B) mRNA,<br>complete cds  | 0.020   | <NONE>  | <NONE>   | <NONE>  |
| 817       | AE001405                              | Plasmodium<br>falciparum<br>chromosome 2,<br>section 42 of 73<br>of the complete<br>sequence  | 0.020   | 2196776   | (AF003342) bunched<br>gene product<br>[Drosophila<br>melanogaster] | 8.4     |
| 818       | AF074387                              | Sambucus nigra<br>hevein-like<br>protein mRNA,<br>complete cds  | 0.020   | 627071  | histidine-rich protein -<br>Plasmodium lophurae                    | 2.8     |
| 819       | Y13304                                | Hylobates<br>hoolock<br>mitochondrial<br>DNA for cytb<br>gene, Horace   | 0.020   | 285580  | (D10043) ORF<br>[Acetobacter<br>pasteurianus]                      | 2.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 820    | Z66539                                | H.sapiens creatine transporter gene                               | 0.020   | 1703594  | (U80439) coded for by C. elegans cDNA yk7c8.5; coded for by C. elegans cDNA yk133b3.5; coded for by C. elegans cDNA yk65a4.5; coded for by C. elegans cDNA yk7c8.3; coded for by C. elegans cDNA CEESQ66F; coded for by C. elegans cDNA yk65a4.3;...   | 0.98    |
| 821    | AF053622                              | Homo sapiens alpha 1,2-mannosidase IB gene, exon 9                | 0.020   | 1352361  | EARLY GROWTH RESPONSE PROTEIN 1 fish >gi 531456 (U12895) egr1 [Danio rerio] rerio]   | 0.72    |
| 822    | M20555                                | Human MHC class II HLA-DRw53-beta (DR4,w4) gene, exons 2,3,4,5,6. | 0.020   | 465569   | HYPOTHETICAL 38.1 KD PROTEIN IN SBCB-HISL INTERGENIC REGION >gi 405956 (U00009) ORF_ID:o349#4; similar to [SwissProt Accession Number P33015] [Escherichia coli] >gi 1736693 gnl PID d1016570 Number P33015] [Escherichia coli] >gi 1788323 (AE000292) putative transport system permease protein [Escherichia coli] | 0.43    |
| 823    | M20555                                | Human MHC class II HLA-DRw53-beta (DR4,w4) gene, exons 2,3,4,5,6. | 0.020   | 1709751  | COENZYME PQQ SYNTHESIS PROTEIN F synthesis F - Pseudomonas fluorescens >gi 929802  | 0.42    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 824    | AJ005015                              | Homo sapiens mRNA for putative SMC-like protein, partial            | 0.020   | 267449   | HYPOTHETICAL 12.5 KD PROTEIN ZK637.2 IN CHROMOSOME III >gi 102507 pir  S1578 7 hypothetical protein 1 (cosmid ZK637) - Caenorhabditis elegans Genefinder; cDNA EST yk217b5.3 comes from this gene; cDNA EST yk217b5.5 comes from this gene; cDNA EST yk340g12.3 comes from this gene; cDNA EST yk340g12.5 comes from this gene; cDNA EST yk428c5.5 co... | 1e-12   |
| 825    | AF034099                              | Laccaria bicolor glyoxal malate synthase protein mRNA, complete cds | 0.020   | 1109847  | (U41538) No definition line found [Caenorhabditis elegans]   | 1e-22   |
| 826    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                            | 0.019   | 132836   | 60S RIBOSOMAL PROTEIN L28 protein L28 [Rattus norvegicus]  | 5.7     |
| 827    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds                       | 0.019   | 2633401  | (Z99109) similar to DNA exonuclease  | 4.5     |
| 828    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                            | 0.019   | 2492604  | MULTIDRUG RESISTANCE PROTEIN CDR2 albicans]  | 4.4     |



| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|---|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 829       | U67538                                | Methanococcus jannaschii section 80 of 150 of the complete genome         | 0.019   | 1723566   | PUTATIVE GLUCOSYLTRANSFERASE C17C9.07 >gi 1314159 gnl PID e 241760 (Z73099) SPAC17C9.07, putative glucosyl transferase len:501, similar to SW:ALG8_YEAST P40351 glucosyltransferase ALG8 pombe] | 2.7     |
| 830       | U56088                                | Human periodic tryptophan protein 2 (PWP2) gene, exons 3 to 14            | 0.019   | 2144804   | collagen alpha 1(II) chain - bovine   | 0.040   |
| 831       | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 0.018   | 1916976   | (U91682) vitelline membrane protein homolog [Aedes aegypti]   | 7.2     |
| 832       | AF026258                              | Onobrychis viciifolia chalcone synthase (CHS) mRNA, complete cds          | 0.018   | 763076  | (Z48799) ZP3 [Cyprinus carpio] >gi 777724 (L41637) egg membrane protein [Cyprinus carpio]   | 5.2     |
| 833       | U95094                                | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                   | 0.009   | 3955011   | (AJ005438) beta adrenoreceptor B  | 0.60    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 834    | X71603                                | C.jejuni VSI DNA > :: emb A39603 A39603 Sequence 2 from Patent WO9417205 > :: gb I76090 I76090 Sequence 2 from patent US 5691138 | 0.008   | <NONE>   | <NONE>   | <NONE>  |
| 835    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds  | 0.008   | 138116   | HEAD FIBER PROTEIN (LATE PROTEIN GP8.5) >gi 75846 pir  WMBP 8H gene 8.5 protein - phage PZA >gi 216057 (M11813) head fiber protein | 8.1     |
| 836    | X91751                                | Bovine herpesvirus type 1 UL7 gene   | 0.008   | 1711436  | SUPEROXIDE DISMUTASE (FE) 1.15.1.1) (Fe) - Pseudomonas aeruginosa >gi 409767   | 5.9     |
| 837    | M95594                                | Arabidopsis thaliana 1-aminocyclopropane-1-carboxylate synthase (ACS2) gene, complete cds.                                       | 0.008   | 683698   | (Z48229) orf1 gene product [Saccharomyces cerevisiae]  | 1e-06   |
| 838    | U67465                                | Methanococcus jannaschii section 7 of 150 of the complete genome   | 0.008   | 3874664  | (Z68493) predicted using Genefinder  | 1e-07   |
| 839    | X72388                                | B.taurus mRNA for filensin   | 0.008   | 100174   | 1-aminocyclopropane-1-carboxylate synthase   | 7e-09   |
| 840    | U22398                                | Human Cdk-inhibitor p57KIP2 (KIP2) mRNA, complete cds.   | 0.008   | 2228750  | (U93868) RNA polymerase III subunit [Homo sapiens]   | 2e-18   |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|---------------------------------------|-----------|---|---------|---|-------------|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 841                                   | L42546    | Xenopus laevis LIM class homeodomain protein                              | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 842                                   | AF041428  | Homo sapiens ribosomal protein s4 X isoform gene, complete cds            | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 843                                   | AF000227  | Secale cereale omega secalin gene, complete cds                           | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 844                                   | D86254    | Human MHC (HLA) DRB intron 1 DNA, partial sequence                        | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 845                                   | AF012899  | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 846                                   | Y07738    | M.musculus gene for vimentin  | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 847                                   | AJ005813  | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                  | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 848                                   | AF055119  | Homo sapiens alpha-tectorin (TECTA) gene, exon 6                          | 0.007   | <NONE>  | <NONE>      | <NONE>  |
| 849                                   | M61195    | Zucchini 1-aminocyclopropane-1-carboxylate synthase                       | 0.007   | <NONE>  | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |             |         |
|--------|---------------------------------------|--|---------|--|-------------|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION | P VALUE |
| 850    | Y11050                                | Homo sapiens DSG3 gene, partial intron and partial exon 6, 140 bp          | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 851    | X61204                                | M.voltae vhuD, vhuG, vhuA, vhuU & vhuB genes                               | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 852    | AB012105                              | Brassica rapa mRNA for SLG45, complete cds                                 | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 853    | S43882                                | telomere: {minichromosome, repeats} [Trypanosoma brucei, Genomic, 1170 nt] | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 854    | L32674                                | Geomydoecus nadleri mitochondrial cytochrome oxidase I gene, partial cds.  | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 855    | U58732                                | Caenorhabditis elegans cosmid F48D6.                                       | 0.007   | <NONE>   | <NONE>      | <NONE>  |
| 856    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds  | 0.007   | <NONE>   | <NONE>      | <NONE>  |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|---|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 857       | Z35284                                | H.sapiens<br>mRNA for<br>MDR3 P-<br>glycoprotein  | 0.007   | 1730696   | HYPOTHETICAL<br>121.1 KD PROTEIN<br>IN BIO3-HXT17<br>INTERGENIC<br>REGION<br>PRECURSOR<br>YNR067c - yeast<br>(Saccharomyces<br>cerevisiae) | 9.5     |
| 858       | X15217                                | Human sno<br>oncogene mRNA<br>for snoA protein,<br>ski-related  | 0.007   | 902455  | (U24203) membrane<br>protein [Escherichia<br>coli]   | 8.8     |
| 859       | AF027173                              | Arabidopsis<br>thaliana<br>cellulose<br>synthase<br>catalytic subunit<br>(Ath-A) mRNA,<br>complete cds  | 0.007   | 1684636   | (Y09454) ORF3<br>[Lactobacillus casei<br>bacteriophage A2]   | 8.3     |
| 860       | AF012899                              | Sambucus nigra<br>ribosome<br>inactivating<br>protein precursor<br>mRNA,<br>complete cds  | 0.007   | 3878803   | (Z48795) R05H5.7<br>[Caenorhabditis<br>elegans]  | 8.3     |
| 861       | S76317                                | TT9=180-200<br>kda membrane<br>protein<br>scavenger<br>receptor<br>homolog {clone<br>18, intron and<br>flanking exons<br>14 and 15}<br>[sheep, lymph<br>node,<br>lymphocytes,<br>Genomic, 308<br>nt, segment 2 of<br>2] | 0.007   | 294747  | (L08174) ORF2<br>[Romanomermis<br>culicivora]  | 7.4     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 862    | D88084                                | Pedicularis verticillata chloroplast DNA, intergenic region between trnT(UGU) and trnL(UAA)5'exon | 0.007   | 2555187  | (AF026789) vitellogenin [Pimpla nipponica]  | 6.9     |
| 863    | X58869                                | Chicken mRNA for aldehyde dehydrogenase   | 0.007   | 115978   | CD30L RECEPTOR PRECURSOR (LYMPHOCYTE ACTIVATION ANTIGEN   | 6.5     |
| 864    | D87120                                | Homo sapiens mRNA for GS3786, complete cds  | 0.007   | 3879589  |   | 5.1     |
| 865    | X68793                                | H.sapiens gene for antithrombin III   | 0.007   | 2358285  | (AF010403) ALR [Homo sapiens]   | 3.8     |
| 866    | AJ001596                              | Danio rerio mRNA for opioid receptor homologue  | 0.007   | 2507509  | HYPOTHETICAL 29.8 KD PROTEIN IN HOLB-PTSG INTERGENIC REGION >gi 1787342 (AE000210) orf, hypothetical protein [Escherichia coli] protein in holB 3'region . [Escherichia coli] | 1.9     |
| 867    | AF061195                              | Streptomyces albus valine dehydrogenase (Vdh) gene, complete cds                                  | 0.007   | 2088768  | (AF003145) B0414.8 gene product [Caenorhabditis elegans]  | 1.9     |
| 868    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme  | 0.007   | 1710105  | UDP-N-ACETYLGLUCOSAMINE 2-EPIMERASE UDP-N-acetylglucosamine 2-epimerase [Plasmid pWQ799]  | 1.7     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 869    | L03398                                | Zebrafish retinoic acid receptor alpha 2.A  | 0.007   | 2239219  | (Z97210) hypothetical protein  | 0.77    |
| 870    | D63484                                | Human mRNA for KIAA0150 gene, partial cds   | 0.007   | 19917  | (Z14014) Pistil extensin like protein, partial CDS only  | 0.61    |
| 871    | M31483                                | Maize glyceraldehyde 3-phosphate dehydrogenase, 3' end.   | 0.007   | 543068   | mucin, tracheobronchial - dog >gi 402558   | 0.45    |
| 872    | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 0.007   | 2494941  | ALPHA-2B ADRENERGIC RECEPTOR adrenoceptor [Cavia porcellus] >gi 1587159 prf  2206293B adrenoceptor alpha2B [Cavia porcellus] | 0.42    |
| 873    | AF064029                              | Helianthus tuberosus lectin 1 mRNA, complete cds  | 0.007   | 1110587  | (S79410) nuclear localization signals Peptide, 140 aa [Mus sp.]  | 0.26    |
| 874    | X88931                                | H.sapiens PAL2A gene  | 0.007   | 1706176  | CUTINASE TRANSCRIPTION FACTOR 1 ALPHA >gi 1262912 (U51671) cutinase transcription factor 1 [Fusarium solani f. sp. pisi]     | 0.21    |
| 875    | S74155                                | zRAR alpha =retinoic acid receptor alpha [zebrafish, embryos, mRNA, 1773 nt]                          | 0.007   | 2239219  | (Z97210) hypothetical protein  | 0.11    |
| 876    | M74193                                | Petromyzon marinus plasma albumin mRNA, complete cds.   | 0.007   | 730888   | OCTAPEPTIDE-REPEAT PROTEIN T2  | 0.011   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|---|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION                                       | P VALUE |
| 877       | U03673                                | Saccharomyces cerevisiae Spp41p (SPP41) gene, complete cds.                           | 0.007   | 3820885   | (AL033126) 65G3.k [Drosophila melanogaster]       | 0.001   |
| 878       | D37766                                | Homo sapiens mRNA for Laminin-5 beta3 chain, complete cds                             | 0.007   | 1235974   | (X96713) collagen [Globodera pallida]             | 3e-06   |
| 879       | AF022388                              | Caenorhabditis elegans putative transcription factor MAB-3 (mab-3) gene, complete cds | 0.007   | 3747107   | (AF095741) unknown [Rattus norvegicus]            | 5e-09   |
| 880       | U89984                                | Acanthamoeba castellanii transformation-sensitive protein homolog mRNA, complete cds  | 0.007   | 1890281   | (U89984) transformation-sensitive protein homolog | 2e-09   |
| 881       | AB020689                              | Homo sapiens mRNA for KIAA0882 protein, partial cds                                   | 0.007   | 3880809   |   | 1e-23   |
| 882       | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 0.006   | <NONE>  | <NONE>  | <NONE>  |
| 883       | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds  | 0.006   | <NONE>  | <NONE>  | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 884    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 885    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 886    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 887    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds   | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 888    | M80529                                | Rattus norvegicus ceruloplasmin gene, exon 1 and 5' flank                            | 0.006   | <NONE>   | <NONE>  | <NONE>  |
| 889    | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 0.006   | 99408  | hypothetical protein 6 - Chlamydomonas reinhardtii transposon >gi 1360717 gnl PID e33461 reinhardtii] | 9.6     |
| 890    | U76523                                | Sambucus nigra lectin precursor mRNA, complete cds                                   | 0.006   | 4039024  | (AF039110) polyprotein [Rubella virus]  | 9.3     |
| 891    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds  | 0.006   | 160533   | (M94428) merozoite surface antigen 1 [Plasmodium vivax]   | 7.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 892    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds   | 0.006   | 4019458  | (AF093984) envelope glycoprotein [Human immunodeficiency virus type 1]   | 7.0     |
| 893    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 0.006   | 1916976  | (U91682) vitelline membrane protein homolog [Aedes aegypti]  | 6.8     |
| 894    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds  | 0.006   | 102059   | promastigote surface antigen-2 (clone 4.6) - Leishmania major (fragment) >gi 9583 (X57135) surface antigen P2 [Leishmania major] | 2.4     |
| 895    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds  | 0.006   | 3171241  | (AF067204) transcription factor BF-1 [Danio rerio]   | 1.0     |
| 896    | X99384                                | M.musculus mRNA for paladin gene   | 0.003   | <NONE>   | <NONE>   | <NONE>  |
| 897    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 0.003   | <NONE>   | <NONE>   | <NONE>  |
| 898    | AE001148                              | Borrelia burgdorferi (section 34 of 70) of the complete genome                       | 0.003   | 4160388  | (AJ011856) ORF Q0255 [Saccharomyces cerevisiae]  | 7.6     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 899    | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds               | 0.003   | 1709213  | NUCLEAR ENVELOPE PORE MEMBRANE PROTEIN POM 121 (PORE MEMBRANE PROTEIN OF 121 KD) (P145) | 1.5     |
| 900    | U72396                                | Lycopersicon esculentum class II small heat shock protein Le-HSP 17.6 mRNA, complete cds           | 0.002   | <NONE>   | <NONE>  | <NONE>  |
| 901    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 0.002   | <NONE>   | <NONE>  | <NONE>  |
| 902    | AF104631                              | Chlamydomonas reinhardtii light harvesting complex II protein precursor (Lhcb3) mRNA, complete cds | 0.002   | <NONE>   | <NONE>  | <NONE>  |
| 903    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 0.002   | <NONE>   | <NONE>  | <NONE>  |
| 904    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds   | 0.002   | <NONE>   | <NONE>  | <NONE>  |
| 905    | M21339                                | Human non-histone chromosomal protein HMG-14 gene, complete cds.                                   | 0.002   | <NONE>   | <NONE>  | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|---------------------------------------|-----------|---|---------|---|-------------|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 906                                   | AF012899  | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds             | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 907                                   | X57103    | Human h-lys gene for lysozyme (upstream region)                                       | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 908                                   | AF074386  | Sambucus nigra hevein-like protein mRNA, complete cds                                 | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 909                                   | U01066    | Human CD4 promoter, partial sequence.   | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 910                                   | L28094    | Barley mRNA sequence.   | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 911                                   | AD000833  | Homo sapiens DNA from chromosome 19-cosmid f19399 (~17 kb EcoRI restriction fragment) | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 912                                   | AJ011701  | Homo sapiens TRHR gene promoter and exons 1-2, partial                                | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 913                                   | AF100694  | Mus musculus Pontin52 mRNA, complete cds  | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 914                                   | AF037062  | Homo sapiens retinol dehydrogenase gene, complete cds                                 | 0.002   | <NONE>  | <NONE>      | <NONE>  |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|---------------------------------------|-----------|--|---------|---|-------------|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 915                                   | AF093268  | Rattus norvegicus homer-1c mRNA, complete cds  | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 916                                   | U67608    | Methanococcus jannaschii section 150 of 150 of the complete genome                   | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 917                                   | AF027173  | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 918                                   | Z46736    | H.sapiens DNA for repeat region (ABM-C82)  | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 919                                   | AB012106  | Brassica rapa mRNA for SRK45, complete cds   | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 920                                   | Z85983    | X.laevis mRNA for NOVA protein   | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 921                                   | AF027173  | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 0.002   | <NONE>  | <NONE>      | <NONE>  |
| 922                                   | S61977    | medium-chain acyl-CoA dehydrogenase {exon 10, intron 10} [human, Genomic, 1407 nt]   | 0.002   | <NONE>  | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 923    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme                             | 0.002   | <NONE>   | <NONE>   | <NONE>  |
| 924    | AB012105                              | Brassica rapa mRNA for SLG45, complete cds   | 0.002   | <NONE>   | <NONE>   | <NONE>  |
| 925    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds   | 0.002   | <NONE>   | <NONE>   | <NONE>  |
| 926    | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 0.002   | <NONE>   | <NONE>   | <NONE>  |
| 927    | X51646                                | H.sapiens DNA for dopamine D2 receptor gene  | 0.002   | 3329125  | (AE001337) Yop C/Gen Secretion Protein D [Chlamydia trachomatis]   | 9.5     |
| 928    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 0.002   | 465762   | HYPOTHETICAL 112.1 KD PROTEIN C06G4.1 IN CHROMOSOME III >gi 630524 pir  S44748 C06G4.1 protein - Caenorhabditis elegans >gi 409292 (L25598) homology with vigilin; coded for by C. elegans cDNA GenBank:M88954 (CEL12C9); putative [Caenorhabditis | 8.9     |
| 929    | U48478                                | Human skeletal muscle ryanodine receptor gene  | 0.002   | 2137221  | co-repressor protein - mouse >gi 642619  | 6.9     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 930    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 0.002   | 806536   | (Z22520) membrane protein [Bacillus acidopullulyticus]   | 6.3     |
| 931    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 0.002   | 3881055  | (AL023844) Y48A6B.1 [Caenorhabditis elegans]   | 5.8     |
| 932    | AF090115                              | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 0.002   | 3878330  | (Z81097) K07A1.4 [Caenorhabditis elegans]  | 4.8     |
| 933    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds   | 0.002   | 137640   | REPLICATION PROTEIN E1 papillomavirus  | 4.0     |
| 934    | AF019660                              | Mus musculus nuclear orphan receptor RORgamma   | 0.002   | 1330365  | (U58757) similar to nucleotide pyrophosphatases  | 3.9     |
| 935    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 0.002   | 1785972  | (U46951) ORF5; Method: conceptual translation supplied by author   | 3.7     |
| 936    | V00508                                | Human gene for epsilon-globin.  | 0.002   | 1333804  | (X56082) protease [Ruminococcus flavefaciens]  | 3.5     |
| 937    | AB012105                              | Brassica rapa mRNA for SLG45, complete cds  | 0.002   | 4153876  | (AC005531) similar to mouse homeodomain-interacting protein kinase 2; similar to AF077659 (PID:g3702958) | 3.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 938    | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme       | 0.002   | 1070461  | ornithine carbamoyltransferase (EC 2.1.3.3) - yeast (Saccharomyces cerevisiae) >gi 929866 (X83502) pid:e130025 [Saccharomyces cerevisiae] >gi 1008256 | 2.8     |
| 939    | S41458                                | rod cGMP phosphodiesterase beta-subunit [human, mRNA, 3231 nt] | 0.002   | 3450883  | (AF083334) fibroin [Antheraea pernyi]   | 1.6     |



| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|---|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 940       | X06286                                | Drosophila melanogaster Gart locus with genes for GARS=phosphoribosylamineglycine ligase, AIRS=phosphoribosylformylglycinamide cycloligase, GART=glycinamide ribotide transformylase > :: gb J02527 DRO GART D.melanogaster Gart gene encoding two polypeptides with GAR synthase, AIR synthase, and GAR transformylase enzyme activities and a pupal cuticle gene nested within intron A of the Gart gene. | 0.002   | 2662054   | (AB004651) isocitrate lyase  | 1.5     |
| 941       | AF015812                              | Homo sapiens RNA helicase p68 (HUMP68) gene, complete cds   | 0.002   | 3641659   | (AB008374) alpha 3 type I collagen                                       | 1.1     |
| 942       | X78925                                | H.sapiens HZF2 mRNA for zinc finger protein   | 0.002   | 141624  | ZINC FINGER PROTEIN ZFP-37 (MALE GERM CELL SPECIFIC ZINC FINGER PROTEIN) | 1.0     |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 943                                   | AF074386  | Sambucus nigra hevein-like protein mRNA, complete cds   | 0.002   | 3879997   | (Z49071) weak similarity with mu-type opioid receptor (Swiss Prot accession number (P33535)) | 1.0     |
| 944                                   | Z69639    | Human DNA sequence from cosmid L241B9, Huntington's Disease Region, chromosome 4p16.3 contains polymorphic VNTR pYNZ32. | 0.002   | 3523162   | (AF076292) TGF-beta/activin signal transducer FAST-1p  | 0.81    |
| 945                                   | AF074387  | Sambucus nigra hevein-like protein mRNA, complete cds   | 0.002   | 2984161   | (AE000761) hypothetical protein [Aquifex aeolicus]   | 0.80    |
| 946                                   | AF093268  | Rattus norvegicus homer-1c mRNA, complete cds   | 0.002   | 101830  | hypothetical protein B - chestnut blight fungus  | 0.72    |
| 947                                   | AF017307  | Homo sapiens Ets-related transcription factor (ERT) mRNA, complete cds  | 0.002   | 200531  | (M18071) prion protein [Mus musculus]  | 0.72    |
| 948                                   | U11383    | Drosophila melanogaster Ovo-1028aa (ovo) mRNA, complete cds.  | 0.002   | 2465207   | (AF016045) OVO-like 1 binding protein [Homo sapiens]   | 0.35    |
| 949                                   | AF012899  | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds   | 0.002   | 3834294   | (U80846) No definition line found [Caenorhabditis elegans]                                   | 0.29    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 950    | AF086315                              | Homo sapiens full length insert cDNA clone ZD52F10                                       | 0.002   | 545067   | (S68356) action potential broadening potassium channel=Shab [Aplysia, bag cell neurons, head ganglia, Peptide, 905 aa] [Aplysia] >gi 743110 prf 201137 5A K channel [Aplysia californica] | 0.15    |
| 951    | X53096                                | S.aureus genes encoding Sau96I DNA methyltransferase and Sau96I restriction endonuclease | 0.002   | 2529575  | (AF018164) kinesin-like protein 3C [Homo sapiens]   | 0.11    |
| 952    | AB012105                              | Brassica rapa mRNA for SLG45, complete cds   | 0.002   | 729918   | LA PROTEIN HOMOLOG (LA RIBONUCLEOPROTEIN) (LA AUTOANTIGEN HOMOLOG)  | 0.092   |
| 953    | X73973                                | G.gallus RAR-gamma2 mRNA for retinoic acid receptor                                      | 0.002   | 586122   | TRICHOHYALIN >gi 423321 pir A40691 trichohyalin - sheep >gi 295941 (Z18361) trichohyalin  | 0.073   |
| 954    | S41458                                | rod cGMP phosphodiesterase beta-subunit [human, mRNA, 3231 nt]                           | 0.002   | 1017427  | (X90569) elastic titin [Homo sapiens]   | 0.013   |
| 955    | M35887                                | D.melanogaster defective chorion-1 fc125 (dec-1) gene, complete cds.                     | 0.002   | 1825606  | (U88169) similar to molybdopterin biosynthesis MOEB proteins [Caenorhabditis elegans]   | 0.008   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 956    | AF034099                              | Laccaria bicolor glyoxal malate synthase protein mRNA, complete cds              | 0.002   | 1825593  | (U88167) D2092.2 gene product [Caenorhabditis elegans] | 1e-06   |
| 957    | AF033929                              | Bactrocera dorsalis strain Tahiti mitochondrial D-loop region, complete sequence | 9e-04   | <NONE>   | <NONE>   | <NONE>  |
| 958    | AB012106                              | Brassica rapa mRNA for SRK45, complete cds                                       | 8e-04   | <NONE>   | <NONE>   | <NONE>  |
| 959    | AF029062                              | Homo sapiens DEAD-box protein (BAT1) gene, partial cds                           | 8e-04   | <NONE>   | <NONE>   | <NONE>  |
| 960    | U70671                                | Human ataxin-2 related protein mRNA, partial cds                                 | 8e-04   | <NONE>   | <NONE>   | <NONE>  |
| 961    | AF051709                              | Dendrocopos leucopterus clone 2 microsatellite HrU2 repeat region                | 8e-04   | <NONE>   | <NONE>   | <NONE>  |
| 962    | X14077                                | Pea phy gene for phytochrome apoprotein  | 8e-04   | <NONE>   | <NONE>   | <NONE>  |
| 963    | AC004497                              | Homo sapiens chromosome 21, P1 clone LBNL#6                                      | 8e-04   | 457146   | (L27838) rhostry protein [Plasmodium yoelii]           | 9.6     |
| 964    | AF077344                              | Homo sapiens cartilage-derived C-type lectin                                     | 8e-04   | 3702123  | (AJ011707) TraD protein [Escherichia coli]             | 8.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 965    | X85117                                | H.sapiens epb72 gene exons 2,3,4,5,6,7  | 8e-04   | 2570059  | (AJ004687) N-4 cytosine-specific Methyltransferase [Neisseria gonorrhoeae]   | 6.8     |
| 966    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 8e-04   | 1345859  | COPPER TRANSPORT PROTEIN CTR1 transport protein - yeast (Saccharomyces cerevisiae) gene product [Saccharomyces cerevisiae] | 6.7     |
| 967    | AF031403                              | Homo sapiens MLL/AF4 translocation breakpoint t(4;11)(q21;23)                         | 8e-04   | 2498926  | SMALL PROTEIN B HOMOLOG A43259, from E. hirae [Mycoplasma pneumoniae]  | 6.6     |
| 968    | L29252                                | Human (clone D13-2) L-iditol-2 dehydrogenase gene, exon 4, exon 5, exon 6 and exon 7. | 8e-04   | 1488070  | (U63997) putative transposase [Enterococcus faecium]   | 5.2     |
| 969    | X16995                                | Mouse N10 gene for a nuclear hormonal binding receptor                                | 8e-04   | 1493833  | (U47323) stromal cell protein [Mus musculus]   | 3.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 970    | M99412                                | Human interleukin-8 receptor (IL8RB) gene, complete cds                        | 8e-04   | 1346101  | 4-AMINOBUTYRATE AMINOTRANSFERASE (GABA AMINOTRANSFERASE) homolog - smut fungus (Ustilago maydis) >gi 881562 Emericella nidulans gamma-amino-n-butyrate transaminase Swiss-Prot Accession Number P14010 [Ustilago maydis]        | 0.83    |
| 971    | U37452                                | Human Down Syndrome region of chromosome 21 genomic sequence, clone A31D6-1C5. | 8e-04   | 4164069  | (AF111093) latrophilin 3 splice variant bbah [Bos taurus]   | 0.26    |
| 972    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                       | 8e-04   | 1352877  | HYPOTHETICAL 13.0 KD PROTEIN IN RAD26-GEF1 INTERGENIC REGION >gi 1077881 pir  S570 57 probable membrane protein YJR038c - yeast (Saccharomyces cerevisiae) >gi 1015688 (Z49538) ORF YJR038c putative [Saccharomyces cerevisiae] | 0.23    |
| 973    | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds                                  | 8e-04   | 1788557  | (AE000312) orf, hypothetical protein [Escherichia coli]   | 0.19    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 974    | X83872                                | H.vulgaris mRNA for cAMP response element binding protein                           | 8e-04   | 1175386  | HYPOTHETICAL 37.7 KD PROTEIN C18B11.06 IN CHROMOSOME I >gi 2130289 pir  S58305 hypothetical protein SPAC18B11.06 - fission yeast hypothetical protein [Schizosaccharomyces pombe] | 0.005   |
| 975    | M32514                                | Rat simple sequence DNA, clone 5.   | 8e-04   | 2394492  | (AF024502) No definition line found [Caenorhabditis elegans]  | 0.002   |
| 976    | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds                               | 8e-04   | 2981631  | (AB012223) ORF2 [Canis familiaris]  | 0.001   |
| 977    | X89211                                | H.sapiens DNA for endogenous retroviral like element                                | 8e-04   | 2065210  | (Y12713) Pro-Pol-dUTPase polyprotein  | 3e-04   |
| 978    | U14391                                | Human myosin-IC mRNA, complete cds.   | 8e-04   | 3142302  | (AC002411) Strong similarity to myosin heavy chain gb Z34293 from A. thaliana. [Arabidopsis thaliana]   | 4e-16   |
| 979    | L13612                                | Drosophila melanogaster dead-box protein D.melanogaster DEAD-box gene, complete CDS | 8e-04   | 3776027  | (AJ010475) RNA helicase [Arabidopsis thaliana]  | 9e-24   |
| 980    | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds                               | 7e-04   | <NONE>   | <NONE>  | <NONE>  |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|-----------|---------------------------------------|---|---------|---|-------------|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 981       | AF100694                              | Mus musculus<br>Pontin52<br>mRNA,<br>complete cds   | 7e-04   | <NONE>  | <NONE>      | <NONE>  |
| 982       | AF093268                              | Rattus<br>norvegicus<br>homer-1c<br>mRNA,<br>complete cds   | 7e-04   | <NONE>  | <NONE>      | <NONE>  |
| 983       | Z73987                                | Human DNA<br>sequence from<br>cosmid N120B6<br>on chromosome<br>22 Contains<br>ESTs, complete<br>sequence [Homo<br>sapiens] | 7e-04   | <NONE>  | <NONE>      | <NONE>  |
| 984       | AB012106                              | Brassica rapa<br>mRNA for<br>SRK45,<br>complete cds   | 7e-04   | <NONE>  | <NONE>      | <NONE>  |
| 985       | AF093268                              | Rattus<br>norvegicus<br>homer-1c<br>mRNA,<br>complete cds   | 7e-04   | <NONE>  | <NONE>      | <NONE>  |
| 986       | AF027174                              | Arabidopsis<br>thaliana<br>cellulose<br>synthase<br>catalytic subunit<br>(Ath-B) mRNA,<br>complete cds                      | 7e-04   | <NONE>  | <NONE>      | <NONE>  |
| 987       | AF100694                              | Mus musculus<br>Pontin52<br>mRNA,<br>complete cds   | 7e-04   | <NONE>  | <NONE>      | <NONE>  |
| 988       | AJ005813                              | Arabidopsis<br>thaliana mRNA<br>for neoxanthin<br>cleavage enzyme   | 7e-04   | <NONE>  | <NONE>      | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION                                      | P VALUE |
| 989    | AF064029                              | Helianthus tuberosus lectin 1 mRNA, complete cds                                     | 7e-04   | <NONE>   | <NONE>   | <NONE>  |
| 990    | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds | 7e-04   | <NONE>   | <NONE>   | <NONE>  |
| 991    | AF027173                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 7e-04   | <NONE>   | <NONE>   | <NONE>  |
| 992    | AF064029                              | Helianthus tuberosus lectin 1 mRNA, complete cds                                     | 7e-04   | <NONE>   | <NONE>   | <NONE>  |
| 993    | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 7e-04   | <NONE>   | <NONE>   | <NONE>  |
| 994    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 7e-04   | 3327230  | (AB014608)<br>KIAA0708 protein<br>[Homo sapiens] | 9.5     |
| 995    | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 7e-04   | 3327230  | (AB014608)<br>KIAA0708 protein<br>[Homo sapiens] | 9.3     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 996                                   | AF074387  | Sambucus nigra hevein-like protein mRNA, complete cds                                | 7e-04   | 3876455   | (Z93380) predicted using Genefinder; similar to 7tm receptor protein [Caenorhabditis elegans]   | 7.1     |
| 997                                   | U76524    | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 7e-04   | 2128771   | hypothetical protein MJ1293 - Methanococcus jannaschii >gi 1591931 (U67570) M. jannaschii predicted coding region MJ1293 [Methanococcus jannaschii] | 6.2     |
| 998                                   | U09412    | Human zinc finger protein ZNF134 mRNA, complete cds                                  | 7e-04   | 1083336   | glutathione transferase (EC 2.5.1.18) piA - mouse   | 5.4     |
| 999                                   | AF027173  | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 7e-04   | 473515  | (M17619) NADH dehydrogenase subunit ND4 [Asterina pectinifera]  | 3.7     |
| 1000                                  | AF012899  | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds            | 7e-04   | 1724097   | (U79772) female sex protein [Mercurialis annua]   | 3.3     |
| 1001                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds   | 7e-04   | 1197103   | (D49747) core, env, and part of E2/NS1  | 3.2     |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1002      | X16995                                | Mouse N10 gene<br>for a nuclear<br>hormonal<br>binding receptor  | 7e-04   | 345372  | unc-5 protein, long<br>form - Caenorhabditis<br>elegans<br>>gi 258529 bbs 118648<br>(S47168) UNC-<br>5=immunoglobulin<br>and thrombospondin<br>type 1 transmembrane<br>protein {alternatively<br>spliced} aa]<br>[Caenorhabditis<br>elegans] >gi 2662596<br>(AF036698) C.<br>elegans UNC-5<br>(NID:g25852) | 2.7     |
| 1003      | U76524                                | Sambucus nigra<br>ribosome<br>inactivating<br>protein precursor<br>mRNA,<br>complete cds               | 7e-04   | 4204220   | (AB022866)<br>mobilization protein   | 2.5     |
| 1004      | AF093268                              | Rattus<br>norvegicus<br>homer-1c<br>mRNA,<br>complete cds  | 7e-04   | 3201550   | (Y17116) fibrinogen-<br>binding protein  | 2.4     |
| 1005      | AF074386                              | Sambucus nigra<br>hevein-like<br>protein mRNA,<br>complete cds   | 7e-04   | 1174264   | (U45966) polypeptide<br>[Hepatitis G virus]  | 0.73    |
| 1006      | AF027173                              | Arabidopsis<br>thaliana<br>cellulose<br>synthase<br>catalytic subunit<br>(Ath-A) mRNA,<br>complete cds | 7e-04   | 135308  | TRANSCRIPTION<br>FACTOR JUN-D  | 0.065   |
| 1007      | X98745                                | H.sapiens EWS<br>gene, intron 6,<br>polymorphism   | 7e-04   | 728836  | !!!! ALU<br>SUBFAMILY SP<br>WARNING ENTRY  | 0.001   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1008   | AJ005813                              | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme   | 7e-04   | 1633564  | (U47924) C8 [Homo sapiens]                               | 9e-09   |
| 1009   | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds  | 6e-04   | 284171   | Ig epsilon chain C region form 3 - human                 | 1.3     |
| 1010   | AB012106                              | Brassica rapa mRNA for SRK45, complete cds   | 6e-04   | 3845262  | (AE001414) BRAHMA ortholog (DNA helicase superfamily II) | 0.25    |
| 1011   | AL034404                              | Human DNA sequence from clone 417C12 on chromosome Xp22.1122.2, complete sequence [Homo sapiens]               | 3e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1012   | M99701                                | Homo sapiens (pp21) mRNA, complete cds.  | 3e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1013   | U00227                                | Ovis aries Merino breed DR beta-chain antigen binding domain, MHC class II DRB (Ovar-DRB24) gene, partial cds. | 3e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1014   | AF074387                              | Sambucus nigra hevein-like protein mRNA, complete cds  | 3e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1015   | U95102                                | Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds  | 3e-04   | 999418   | (L19655) ORF [Tomato ringspot virus]                     | 8.3     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1016                                  | AB012106  | Brassica rapa mRNA for SRK45, complete cds   | 3e-04   | 2367460   | (AF011415) putative pheromone receptor [Mus musculus]  | 7.0     |
| 1017                                  | AJ010737  | Mus musculus DNA for microsatellite 3kb upstream lbp gene                            | 3e-04   | 4106549   | (AF104411) neuronal-specific septin 3 [Mus musculus]   | 5.5     |
| 1018                                  | AF053137  | Homo sapiens histone deacetylase 3 gene, exons 4, 5, 6, 7, 8, 9, and 10              | 3e-04   | 416702  | NADH-DEPENDENT FLAVIN OXIDOREDUCTASE acid-inducible - Eubacterium sp >gi 1381570 (U57489) NADH:flavin oxidoreductase [Eubacterium sp. VPI 12708] | 5.3     |
| 1019                                  | AF027173  | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds | 3e-04   | 1785789   | (Y08502) orf111d [Arabidopsis thaliana]  | 5.1     |
| 1020                                  | AC004173  | Homo sapiens clone UWGC:y23x011 from 6p21, complete sequence [Homo sapiens]          | 3e-04   | 558521  | (D28917) polyprotein [Hepatitis C virus]   | 1.1     |
| 1021                                  | X57025    | Human IGF-I mRNA for insulin-like growth factor I                                    | 3e-04   | 4206707   | (AF118122) putative outer membrane protein OmpU  | 0.65    |
| 1022                                  | X77090    | H.sapiens IL-1Ra gene.   | 3e-04   | 1065941   | (U40799) F42C5.7 gene product [Caenorhabditis elegans]   | 0.12    |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|---|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1023                                  | M34651    | Pseudorabies virus with upstream and downstream sequences.                                    | 3e-04   | 2746853   | (AF040650) contains similarity to sodium-potassium-chloride cotransport proteins  | 7e-05   |
| 1024                                  | Z36011    | S.cerevisiae chromosome II reading frame ORF YBR142w  | 3e-04   | 2500537   | PROBABLE ATP-DEPENDENT RNA HELICASE HAS1<br>>gi 626265 pir  S4745<br>1 hypothetical protein YMR290c RNA helicase<br>[Saccharomyces cerevisiae]  | 4e-08   |
| 1025                                  | AF020286  | Dictyostelium discoideum 2034 gene, partial cds   | 3e-04   | 1465834   | (U64857) No definition line found<br>[Caenorhabditis elegans]   | 6e-14   |
| 1026                                  | L26049    | Chlamydomonas reinhardtii dynein heavy chain alpha (ODA11) gene, exons 2-15, and partial cds. | 3e-04   | 3876775   | (Z81077) predicted using Genefinder; Similarity to Yeast protein 8248 (TR:G587531)  | 9e-15   |
| 1027                                  | AF020286  | Dictyostelium discoideum 2034 gene, partial cds   | 3e-04   | 1465834   | (U64857) No definition line found<br>[Caenorhabditis elegans]   | 1e-17   |
| 1028                                  | X79811    | S.cerevisiae ACT3 gene  | 3e-04   | 3876090   | (Z69635) Similarity to Yeast uridine kinase (SW:URK1_YEAST); cDNA EST EMBL:Z14695 comes from this gene; cDNA EST CEMSE17F comes from this gene; cDNA EST EMBL:D67355 comes from this gene; cDNA EST yk209h1.5 comes from this ge... | 7e-31   |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |             |         |
|---------------------------------------|-----------|--|---------|---|-------------|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION | P VALUE |
| 1029                                  | AF027173  | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-A) mRNA, complete cds                 | 2e-04   | <NONE>  | <NONE>      | <NONE>  |
| 1030                                  | M22970    | Human pancreatic phospholipase A-2 (PLA-2) gene, exons 1 to 3.                                       | 2e-04   | <NONE>  | <NONE>      | <NONE>  |
| 1031                                  | Z68686    | Human DNA sequence from cosmid N2E9 on chromosome 22. Contains EST, complete sequence [Homo sapiens] | 2e-04   | <NONE>  | <NONE>      | <NONE>  |
| 1032                                  | X95154    | H.sapiens brca2 gene exon 4 > :: emb A62779 A62779 Sequence 20 from Patent WO9719110                 | 2e-04   | <NONE>  | <NONE>      | <NONE>  |
| 1033                                  | AJ005813  | Arabidopsis thaliana mRNA for neoxanthin cleavage enzyme   | 2e-04   | <NONE>  | <NONE>      | <NONE>  |
| 1034                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds   | 2e-04   | <NONE>  | <NONE>      | <NONE>  |
| 1035                                  | AE001415  | Plasmodium falciparum chromosome 2, section 52 of 73 of the complete sequence                        | 2e-04   | <NONE>  | <NONE>      | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1036   | AF090115                              | Lycopersicon<br>esculentum<br>cytosolic class II<br>small heat shock<br>protein HCT2<br>(HSP17.4)<br>mRNA,<br>complete cds | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1037   | AC000958                              | Homo sapiens<br>(subclone 6_d9<br>from P1 H21)<br>DNA sequence   | 2e-04   | <NONE>   | <NONE>   | <NONE>  |
| 1038   | AF093268                              | Rattus<br>norvegicus<br>homer-1c<br>mRNA,<br>complete cds  | 2e-04   | 2501523  | CD59<br>GLYCOPROTEIN<br>PRECURSOR  | 7.1     |
| 1039   | U76524                                | Sambucus nigra<br>ribosome<br>inactivating<br>protein precursor<br>mRNA,<br>complete cds                                   | 2e-04   | 2765360  | (Y13925) cathepsin L2<br>[Penaeus vannamei]  | 6.8     |
| 1040   | AF027174                              | Arabidopsis<br>thaliana<br>cellulose<br>synthase<br>catalytic subunit<br>(Ath-B) mRNA,<br>complete cds                     | 2e-04   | 133636   | RNA POLYMERASE<br>>gi 67126 pir  RRXPL<br>C RNA-directed RNA<br>polymerase (EC<br>2.7.7.48) -<br>lymphocytic<br>choriomeningitis virus<br>(strain Armstrong 53b)<br>>gi 331369 | 5.2     |
| 1041   | AB012106                              | Brassica rapa<br>mRNA for<br>SRK45,<br>complete cds  | 2e-04   | 3822155  | (AF074613) type II<br>secretion protein<br>[Escherichia coli<br>O157:H7]   | 4.0     |
| 1042   | U76524                                | Sambucus nigra<br>ribosome<br>inactivating<br>protein precursor<br>mRNA,<br>complete cds                                   | 2e-04   | 1718125  | REGULATORY<br>PROTEIN E2<br>>gi 1020222 type 36]   | 0.38    |



| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1043                                  | X17058    | Sus scrofa<br>mRNA for<br>glucose<br>transporte<br>protein                 | 2e-04   | 3341906   | (AB009593) xylose<br>transporter                            | 2e-15   |
| 1044                                  | AF008216  | Homo sapiens<br>candidate tumor<br>suppressor<br>pp32r1                    | 1e-04   | <NONE>  | <NONE>  | <NONE>  |
| 1045                                  | X98890    | S.tuberosum<br>mRNA for<br>inorganic<br>phosphate<br>transporter,<br>StPT1 | 1e-04   | 624126  | (U42580) a65L<br>[Paramecium bursaria<br>Chlorella virus 1] | 7.9     |
| 1046                                  | L14930    | Glycine max<br>(Rab7p) mRNA,<br>complete cds.                              | 9e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1047                                  | AJ009970  | Mus musculus<br>thromboxane A2<br>receptor gene,<br>exon 3, partial        | 9e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1048                                  | Y11896    | M.musculus<br>mRNA for Brx<br>gene, partial                                | 9e-05   | <NONE>  | <NONE>  | <NONE>  |
| 1049                                  | L10832    | Polistes<br>annularis (clone<br>pan48AAT)<br>tandem repeat<br>region.      | 9e-05   | <NONE>  | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1050   | AF055011                              | Homo sapiens clone 24587 mRNA sequence                                    | 9e-05   | 3880586  | (Z79758) cDNA EST EMBL:D28009 comes from this gene; cDNA EST EMBL:D28008 comes from this gene; cDNA EST EMBL:D32478 comes from this gene; cDNA EST EMBL:D34508 comes from this gene; cDNA EST EMBL:D37581 comes from this gene; ... | 7.6     |
| 1051   | U76524                                | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 9e-05   | 3024292  | RHODOPSIN >gi 2290717 (AF000947) rhodopsin [Sepia officinalis]  | 6.7     |
| 1052   | Z58294                                | H.sapiens CpG DNA, clone 34d6, forward read cpg34d6.ftla.                 | 9e-05   | 3885496  | (AF064825) heparin/heparan sulfate N-acetylglucosaminyl N-deacetylase/N-sulfotransferase [Bos taurus]   | 0.65    |
| 1053   | D87451                                | Human mRNA for KIAA0262 gene, complete cds                                | 9e-05   | 3874739  | (Z66495) similar to claustrin like  | 0.004   |
| 1054   | L37092                                | Mus musculus cyclin-dependent kinase homologue                            | 9e-05   | 3080513  | (AL022598) hypothetical protein   | 4e-09   |
| 1055   | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds                     | 8e-05   | <NONE>   | <NONE>  | <NONE>  |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|---|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1056      | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds    | 8e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1057      | AF074386                              | Sambucus nigra hevein-like protein mRNA, complete cds                                   | 8e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1058      | D10102                                | Homo sapiens DNA from cosmid clone:844, GT repeat sequence                              | 8e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1059      | U72396                                | Lycopersicon esculentum class II small heat shock protein Le-HSP17.6 mRNA, complete cds | 8e-05   | 1176475   | HYPOTHETICAL 80.4 KD PROTEIN IN SMC3-MRPL8 INTERGENIC REGION<br>>gi 1078237 pir  S568 49 probable membrane protein YJL073w - yeast (Saccharomyces cerevisiae) >gi895898 (X88851) hypothetical protein YJL073w [Saccharomyces cerevisiae] | 6.0     |
| 1060      | X71934                                | H.sapiens XB gene for tenascin-X, repeat XIII   | 8e-05   | 285207  | microtubule-associated protein, 110K tau - rat >gi 207158 (M84156) big tau [Rattus norvegicus]   | 3.7     |
| 1061      | AF027174                              | Arabidopsis thaliana cellulose synthase catalytic subunit (Ath-B) mRNA, complete cds    | 8e-05   | 4049682   | (AF063866) ORF MSV092 hypothetical protein [Melanoplus sanguinipes entomopoxvirus]   | 2.1     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1062                                  | AF090115  | Lycopersicon<br>esculentum<br>cytosolic class II<br>small heat shock<br>protein HCT2<br>(HSP17.4)<br>mRNA,<br>complete cds | 8e-05   | 3861019   | (AJ235271) unknown<br>[Rickettsia<br>prowazekii]   | 5e-14   |
| 1063                                  | AF027174  | Arabidopsis<br>thaliana<br>cellulose<br>synthase<br>catalytic subunit<br>(Ath-B) mRNA,<br>complete cds                     | 7e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1064                                  | L04193    | Human lens<br>membrane<br>protein (mp19)<br>gene, exon 11.   | 7e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1065                                  | X61609    | B.napus gene for<br>LHC II Type III<br>chlorophyll a/b<br>binding protein  | 7e-05   | 2132314   | hypothetical protein<br>YPR174c - yeast<br>similarity to a nuclear<br>lamin from C. elegans<br>(PIR accession number<br>S42257)<br>[Saccharomyces<br>cerevisiae] | 8.9     |
| 1066                                  | AF064029  | Helianthus<br>tuberosus lectin<br>1 mRNA,<br>complete cds  | 7e-05   | 2979422   | (AB006757) PCDH7<br>(BH-Pcdh)c [Homo<br>sapiens]   | 5.7     |
| 1067                                  | AF027173  | Arabidopsis<br>thaliana<br>cellulose<br>synthase<br>catalytic subunit<br>(Ath-A) mRNA,<br>complete cds                     | 7e-05   | 2493696   | HYPOTHETICAL<br>21.5 KD PROTEIN<br>(ORF 185)<br>>gi 1480440 (U34204)<br>ORF185; hypothetical<br>21.4 kD protein<br>[Brassica oleracea]                           | 5.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1068   | AF093268                              | Rattus norvegicus homer-1c mRNA, complete cds  | 7e-05   | 2501029  | PROBABLE LEUCYL-TRNA SYNTHETASE, MITOCHONDRIAL PRECURSOR (LEUCINE--TRNA LIGASE) (LEURS) KIAA0028 [Homo sapiens]                                    | 1.4     |
| 1069   | Z68758                                | Human DNA sequence from cosmid cN85E10 on chromosome 22q11.2-qter  | 3e-05   | <NONE>   | <NONE>   | <NONE>  |
| 1070   | X60653                                | human Histone H3.3 pseudogene (CIR-456)  | 3e-05   | <NONE>   | <NONE>   | <NONE>  |
| 1071   | Z58294                                | H.sapiens CpG DNA, clone 34d6, forward read cpg34d6.ftla.  | 3e-05   | 1706241  | GUANYLYL CYCLASE GC-E PRECURSOR cyclase receptor [Mus musculus]  | 9.6     |
| 1072   | AF043251                              | Homo sapiens mitochondrial outer membrane protein (Tom40) gene, nuclear gene encoding mitochondrial protein, exons 1 through 6 | 3e-05   | 113980   | AMINE OXIDASE [FLAVIN-CONTAINING] B oxidase (flavin-containing) (EC 1.4.3.4) B - human B [human, platelet, Peptide Partial, 520 aa] [Homo sapiens] | 8.9     |
| 1073   | M31104                                | Chicken progesterone receptor gene, encoding forms A and B, exons 1 and 2.   | 3e-05   | 1170841  | IG GAMMA LAMBDA CHAIN V-II REGION  | 4.8     |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1074                                  | AF012899  | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                             | 3e-05   | 543684  | ribosomal protein S3 - Chlamydomonas humicola chloroplast (fragment) | 4.2     |
| 1075                                  | L22206    | Human vasopressin receptor V2 gene, complete cds.   | 3e-05   | 791207  | (U20615) Gnot1 homeodomain protein [Gallus gallus]                   | 1.8     |
| 1076                                  | AF093268  | Rattus norvegicus homer-1c mRNA, complete cds   | 3e-05   | 3237340   | (AF033361) polyprotein [Hepatitis C virus]                           | 0.94    |
| 1077                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds  | 3e-05   | 2879805   | (AL021813) hypothetical protein                                      | 0.001   |
| 1078                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds  | 3e-05   | 3877951   | (Z81555) predicted using Genefinder                                  | 3e-07   |
| 1079                                  | AF090115  | Lycopersicon esculentum cytosolic class II small heat shock protein HCT2 (HSP17.4) mRNA, complete cds | 2e-05   | <NONE>  | <NONE>   | <NONE>  |
| 1080                                  | AF064029  | Helianthus tuberosus lectin 1 mRNA, complete cds  | 2e-05   | 3880197   | (Z81132) predicted using Genefinder                                  | 2.4     |
| 1081                                  | AF087989  | Homo sapiens full length insert cDNA clone YX29D10  | 2e-05   | 113667  | !!!! ALU CLASS B WARNING ENTRY !!!!                                  | 1.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1082   | AF064029                              | Helianthus tuberosus lectin 1 mRNA, complete cds                          | 2e-05   | 474896   | (L31967) mating type protein [Coprinus cinereus]                                   | 1.4     |
| 1083   | AF064029                              | Helianthus tuberosus lectin 1 mRNA, complete cds                          | 2e-05   | 2266988  | (Y13274) M33 polycomb-like protein [Mus musculus]                                  | 0.62    |
| 1084   | U67415                                | Equus caballus UCD-E-CA-467 dinucleotide repeat region, complete sequence | 1e-05   | <NONE>   | <NONE>   | <NONE>  |
| 1085   | X67277                                | H.sapiens BGP gene for biliary glycoprotein, promoter region and exon 1   | 1e-05   | <NONE>   | <NONE>   | <NONE>  |
| 1086   | X85117                                | H.sapiens epb72 gene exons 2,3,4,5,6,7                                    | 1e-05   | <NONE>   | <NONE>   | <NONE>  |
| 1087   | U88328                                | Mus musculus suppressor of cytokine signalling-3                          | 1e-05   | 443877   | (Z29457) core region; pid:g443877 [Hepatitis C virus] virus]                       | 3.9     |
| 1088   | Y12853                                | Homo sapiens P2X7 gene, exon 4-8  | 1e-05   | 3878726  | (Z66498) similar to cuticle collagen; cDNA EST EMBL:D75584 comes from this gene    | 0.36    |
| 1089   | AE001140                              | Borrelia burgdorferi (section 26 of 70) of the complete genome            | 1e-05   | 3860719  | (AJ235270) GLUTAMYL-tRNA AMIDOTRANSFERASE SUBUNIT A (gatA) [Rickettsia prowazekii] | 4e-15   |
| 1090   | AJ224112                              | Homo sapiens gamma adaptin gene, exon 2 and flanking intronic sequences   | 9e-06   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1091   | AB000565                              | Homo sapiens DNA for repeat sequence Alu                       | 9e-06   | 72879  | translation initiation factor IF-2 - Escherichia coli   | 5.1     |
| 1092   | Z78985                                | H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA20B4 | 9e-06   | 159975   | (M65164) 51C surface protein [Paramecium tetraurelia]   | 4.8     |
| 1093   | Z21677                                | Thermotoga maritima DNA for spc operon                         | 9e-06   | 585879   | 50S RIBOSOMAL PROTEIN L2 maritima >gi 437926 (Z21677) ribosomal protein L2                        | 7e-14   |
| 1094   | AF031494                              | Drosophila hydei Dhc7 (Threads) mRNA, complete cds             | 9e-06   | 729377   | DYNEIN BETA CHAIN, CILIARY sea urchin (Anthocidaris crassispina) chain [Anthocidaris crassispina] | 4e-18   |
| 1095   | AF051315                              | Homo sapiens placental protein 17a1 (PP17) mRNA, complete cds  | 4e-06   | <NONE>   | <NONE>  | <NONE>  |
| 1096   | AC001460                              | Homo sapiens (subclone 2_f4 from BAC H107) DNA sequence        | 4e-06   | 2648304  | (AE000952) ISA1214-6, putative transposase  | 6.2     |
| 1097   | X85030                                | H.sapiens mRNA for skeletal muscle-specific calpain            | 4e-06   | 4239857  | (AB016726) calpain [Schistosoma japonicum]  | 0.006   |
| 1098   | M75162                                | Human polymorphic arylamine N-acetyltransferase                | 3e-06   | <NONE>   | <NONE>  | <NONE>  |



| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1099                                  | AB009999  | Rattus norvegicus mRNA for CDP-diacylglycerol synthase, complete cds | 3e-06   | 3879045   | (Z70309) R102.6 [Caenorhabditis elegans]   | 7.3     |
| 1100                                  | Z78985    | H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA20B4       | 3e-06   | 266529  | MERCURIC REDUCTASE (HG(II) REDUCTASE) >gi 418744 pir  S30168 mercury(II) reductase | 6.5     |
| 1101                                  | AB012190  | Homo sapiens mRNA for Nedd8-activating enzyme hUba3, complete cds    | 3e-06   | 3877938   | (Z79697) F58H10.1 [Caenorhabditis elegans]   | 6.3     |
| 1102                                  | AF041056  | Homo sapiens WSCR4 gene, exons 3 and 4                               | 3e-06   | 1568583   | (Z80775) hypothetical protein Rv0044c  | 1.9     |
| 1103                                  | X00777    | Mouse E(d) beta gene 5' flanking region and exon 1                   | 3e-06   | 1680722   | (U72497) fatty acid amide hydrolase [Rattus norvegicus]                            | 0.008   |
| 1104                                  | D21205    | Human mRNA for estrogen responsive finger protein, complete cds      | 3e-06   | 563127  | (U09825) acid finger protein [Homo sapiens]  | 1e-05   |
| 1105                                  | Z47046    | Human cosmid QLL2C9 from Xq28  | 1e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1106                                  | L26261    | Human MHC class III HLA-RP1 gene.                                    | 1e-06   | <NONE>  | <NONE>   | <NONE>  |
| 1107                                  | M13402    | Rat 5S RNA gene, clone 5S-2.   | 1e-06   | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1108   | X68793                                | H.sapiens gene for antithrombin III   | 1e-06   | <NONE>   | <NONE>  | <NONE>  |
| 1109   | AF003540                              | Homo sapiens Krueppel family zinc finger protein  | 1e-06   | 2507553  | ZINC FINGER PROTEIN 33A (ZINC FINGER PROTEIN KOX31) (KIAA0065) (HA0946) Kruppel-related. [Homo sapiens]   | 0.098   |
| 1110   | L42096                                | Homo sapiens (subclone 10_d2 from P1 H21) DNA sequence.   | 1e-06   | 1330401  | (U58762) T27F7.1 gene product [Caenorhabditis elegans]  | 0.015   |
| 1111   | Z69925                                | Human DNA sequence from cosmid cN116A5, between markers D22S280 and D22S86 on chromosome 22q12 contains EST | 9e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1112   | D90217                                | S. cerevisiae gene for YmL33, mitochondrial ribosomal proteins of large subunit                             | 9e-07   | 3879097  | (Z81109) predicted using Genefinder; similar to sodium/phosphate transporter; cDNA EST yk326f6.3 comes from this gene; cDNA EST yk326f6.5 comes from this gene [Caenorhabditis elegans] | 7.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1113   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 9e-07   | 1330345  | (U58755) coded for by C. elegans cDNA yk34b1.5; coded for by C. elegans cDNA yk13h10.5; coded for by C. elegans cDNA yk46e8.5; coded for by C. elegans cDNA yk46d5.5; coded for by C. elegans cDNA yk43c2.5; coded for by C. elegans cDNA yk46e8.... | 2e-29   |
| 1114   | AF086562                              | Homo sapiens full length insert cDNA clone ZE16C03                        | 4e-07   | 1072210  | (U40945) coded for by C. elegans cDNA yk74b9.3; coded for by C. elegans cDNA yk74b9.5; similar to repeat of calcium channel alpha subunits; similar to tetracycline resistance protein; similar to hypothetical protein in HSP30-PMP1 region (SP...  | 3.9     |
| 1115   | L39062                                | Homo sapiens interleukin 9 receptor IL9R pseudogene, exons 1-9            | 4e-07   | 3879983  | (Z46795) similar to transforming protein etc2; cDNA EST EMBL:D34137 comes from this gene; cDNA EST EMBL:D37172 comes from this gene; cDNA EST EMBL:D76266 comes from this gene; cDNA EST EMBL:D70493 comes from this gene; cDNA ...                  | 3.3     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1116   | Z69364                                | Human DNA sequence from cosmid L96F8, Huntington's Disease Region, chromosome 4p16.3 contains EST and cDNA.<br>> ::<br>emb Z69365 HS L96F8A Human DNA sequence from cosmid L96F8, Huntington's Disease Region, chromosome 4p16.3 contains EST and cDNA. | 4e-07   | 3493176  | (AF022889) latent TGF beta binding protein [Mus musculus]   | 3.0     |
| 1117   | D79986                                | Human mRNA for KIAA0164 gene, complete cds  | 4e-07   | 4038031  | (AC005936) hypothetical protein [Arabidopsis thaliana]  | 0.30    |
| 1118   | D43950                                | Human mRNA for KIAA0098 gene, partial cds   | 3e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1119   | AF037168                              | Arabidopsis thaliana DnaJ homologue (AtJ6) mRNA, complete cds   | 3e-07   | 3881075  | (AL032657) predicted using Genefinder; similar to DnaJ domain ; Thioredoxin; cDNA EST yk433f3.5 comes from this gene; cDNA EST EMBL:D32359 comes from this gene; cDNA EST EMBL:D34721 comes from this gene; cDNA EST yk433f3.3 c... | 3e-09   |
| 1120   | X69838                                | H.sapiens mRNA for G9a  | 3e-07   | 3873414  | (U00043) similar to D. melanogaster trithorax protein   | 3e-29   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1121   | AB011124                              | Homo sapiens mRNA for KIAA0552 protein, complete cds                           | 2e-07   | 2618749  | (U90880) hypothetical protein 2; predicted using XGrail   | 2.0     |
| 1122   | K03012                                | Human cellular fms proto-oncogene, partial cds.                                | 1e-07   | <NONE>   | <NONE>  | <NONE>  |
| 1123   | AB016195                              | Homo sapiens DNA, microsatellite and Alu repeat region                         | 1e-07   | 728837   | !!!! ALU SUBFAMILY SQ WARNING ENTRY   | 0.095   |
| 1124   | Y16795                                | Homo sapiens psihHaA pseudogene  | 4e-08   | <NONE>   | <NONE>  | <NONE>  |
| 1125   | AB012624                              | Homo sapiens FLI1 gene for ERGB transcription factor, intron 4 and partial cds | 4e-08   | 728836   | !!!! ALU SUBFAMILY SP WARNING ENTRY   | 3.6     |
| 1126   | AJ131341                              | Homo sapiens ogg1 gene, exons 1-7  | 4e-08   | 113668   | !!!! ALU CLASS C WARNING ENTRY<br>!!!!  | 3e-05   |
| 1127   | L81902                                | Homo sapiens (subclone 1_c10 from P1 H69) DNA sequence                         | 3e-08   | 4225950  | (AJ132701) centaurin gamma1B  | 1.8     |
| 1128   | Y17968                                | Gallus gallus mRNA for high mobility group 1 protein                           | 3e-08   | 3041855  | (AC004537) similar to tumor suppressor p33ING1; similar to AF044076 (PID:g2829208) [Homo sapiens] | 3e-31   |
| 1129   | Y13901                                | Homo sapiens FGFR-4 gene   | 1e-08   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1130   | L22024                                | Mesocricetus auratus serum amyloid P component gene, complete cds.                               | 1e-08   | <NONE>   | <NONE>   | <NONE>  |
| 1131   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                        | 1e-08   | <NONE>   | <NONE>   | <NONE>  |
| 1132   | X14034                                | Human mRNA for phospholipase C > :: gb M37238 HU MPL C Human phospholipase C mRNA, complete cds. | 1e-08   | <NONE>   | <NONE>   | <NONE>  |
| 1133   | Z59381                                | H.sapiens CpG DNA, clone 152b10, forward read cpg152b10.f1a.                                     | 1e-08   | <NONE>   | <NONE>   | <NONE>  |
| 1134   | L81839                                | Homo sapiens (subclone 2_h3 from P1 H43) DNA sequence  | 1e-08   | <NONE>   | <NONE>   | <NONE>  |
| 1135   | X14448                                | Human GLA gene for alpha-D-galactosidase A (EC 3.2.1.22)   | 1e-08   | 3334427  | HYPOTHETICAL PROTEIN MJ1207 Methanococcus jannaschii > gi 1591837 (U67562) protease synthase and sporulation negative regulator Pai 1, putative [Methanococcus jannaschii] | 9.1     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1136   | AL023774                              | Human DNA sequence from clone 799F15 on chromosome Xq25, complete sequence [Homo sapiens]          | 1e-08   | 1354935  | (U58330) probable copper-transporting atpase  | 1.2     |
| 1137   | X64639                                | H.sapiens DNA repetitive subtelomeric-like sequence (522 bp)                                       | 1e-08   | 77356  | hypothetical 70K protein - eggplant mosaic virus  | 0.098   |
| 1138   | U97058                                | Human HuD gene, 5'UTR  | 5e-09   | 3387886  | (AF070530) unknown [Homo sapiens]   | 9.5     |
| 1139   | Z82181                                | Human DNA sequence from cosmid E86D10 on chromosome 22. contains ESTs, exontrap, complete sequence | 5e-09   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY  | 8.4     |
| 1140   | AJ006587                              | Mus musculus mRNA for translation initiation factor eIF2 gamma X                                   | 5e-09   | 1872200  | (U22376) alternatively spliced product using exon 13A   | 0.64    |
| 1141   | Y11108                                | H.sapiens WNT8B gene   | 4e-09   | 2854198  | (AF045646) contains similarity to collagens   | 4.0     |
| 1142   | AE001223                              | Treponema pallidum section 39 of 87 of the complete genome   | 4e-09   | 3334189  | CELL DIVISION PROTEIN FTSY HOMOLOG  | 1.5     |
| 1143   | Z47046                                | Human cosmid QLL2C9 from Xq28  | 4e-09   | 104045   | fibroblast growth factor receptor A1 precursor - African clawed frog >gi 214894 (M55163) fibroblast growth factor receptor [Xenopus laevis] | 1.3     |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1144                                  | AG000746  | Homo sapiens genomic DNA, 21q region, clone: T171Bm40                                    | 4e-09   | 113666  | !!!! ALU CLASS A WARNING ENTRY !!!!  | 0.33    |
| 1145                                  | M74002    | Human arginine-rich nuclear protein mRNA, complete cds.                                  | 4e-09   | 3875371   |  | 3e-06   |
| 1146                                  | U95094    | Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds                                  | 2e-09   | 2494337   | ENDO-1,4-BETA-XYLANASE PRECURSOR sp.]  | 4.9     |
| 1147                                  | U20554    | Drosophila melanogaster UDP-glucose:glycoprotein glycosyltransferase mRNA, complete cds. | 2e-09   | 2499087   | UDP-GLUCOSE:GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR (DUGT) glucosyltransferase - fruit fly (Drosophila sp.) glucosyltransferase precursor [Drosophila melanogaster] | 4e-24   |
| 1148                                  | Z56162    | H. sapiens CpG DNA, clone 91c9, forward read cpg91c9.ftla.                               | 1e-09   | <NONE>  | <NONE>   | <NONE>  |
| 1149                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds   | 1e-09   | 1002424   | (U25739) YSPL-1 form 1 [Mus musculus]  | 8.9     |
| 1150                                  | M85276    | Homo sapiens NKG5 gene, complete cds.  | 1e-09   | 2315436   | (AF016447) No definition line found [Caenorhabditis elegans]   | 8.3     |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1151   | M94065                                | Human dihydroorotate dehydrogenase mRNA, 3' end.   | 1e-09   | 3892656  | (AB014464) MGC-24v [Mus musculus]                   | 6.2     |
| 1152   | AJ131895                              | Homo sapiens genomic CAG repeat element, clone 60o2(250)   | 5e-10   | <NONE>   | <NONE>  | <NONE>  |
| 1153   | Z82181                                | Human DNA sequence from cosmid E86D10 on chromosome 22. contains ESTs, exontrap, complete sequence | 5e-10   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY                  | 7.9     |
| 1154   | AJ224442                              | Homo sapiens mRNA for putative methyltransferase   | 5e-10   | 113667   | !!!! ALU CLASS B WARNING ENTRY                      | 0.15    |
| 1155   | AJ010230                              | Homo sapiens RET finger protein-like 1 antisense transcript, partial                               | 5e-10   | 728834   | !!!! ALU SUBFAMILY SB2 WARNING ENTRY                | 0.006   |
| 1156   | AF111116                              | Homo sapiens silencer of death domains (SODD) mRNA, complete cds                                   | 5e-10   | 4160014  | (AF111116) silencer of death domains [Homo sapiens] | 2e-08   |
| 1157   | Z97017                                | Homo sapiens mRNA for hypothetical protein   | 4e-10   | <NONE>   | <NONE>  | <NONE>  |
| 1158   | AF001298                              | Homo sapiens type II integral membrane protein   | 4e-10   | <NONE>   | <NONE>  | <NONE>  |
| 1159   | Y11395                                | H. sapiens mRNA for p40  | 2e-10   | 1000340  | (U34384) CheW [Borrelia burgdorferi]                | 2.4     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1160   | U41096                                | Human non-coding sequence upstream from DOC-2 gene on chromosome 5                             | 2e-10   | 728837   | !!!! ALU SUBFAMILY SQ WARNING ENTRY  | 0.28    |
| 1161   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                      | 6e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1162   | Z36111                                | S. cerevisiae chromosome II reading frame ORF YBR242w  | 6e-11   | 2213560  | (Z97052) hypothetical protein  | 3e-27   |
| 1163   | D89174                                | Schizosaccharomyces pombe mRNA, partial cds, clone: SY 1004                                    | 6e-11   | 3879758  | (Z80220) Similarity to yeast protein TREMBL ID E246895); cDNA EST EMBL:T00018 comes from this gene; cDNA EST EMBL:C13908 comes from this gene; cDNA EST EMBL:C11656 comes from this gene; cDNA EST yk234a5.3 comes from this ge... | 4e-30   |
| 1164   | Z95437                                | Human DNA sequence from cosmid A1 on chromosome 6 contains ESTs. HERV like retroviral sequence | 5e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1165   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds                      | 5e-11   | 3886065  | (AF106581) contains similarity to C4-type zinc fingers   | 4.9     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1166   | X56997                                | Human UbA52 gene coding for ubiquitin-52 amino acid fusion protein                                 | 2e-11   | <NONE>   | <NONE>   | <NONE>  |
| 1167   | AF086253                              | Homo sapiens full length insert cDNA clone ZD40G12   | 2e-11   | 2134780  | apoptosis inhibitor IAP homolog - human  | 3.8     |
| 1168   | AB018314                              | Human sapiens mRNA for KIAA0771 protein, partial cds   | 2e-11   | 3024343  | P53-BINDING PROTEIN 53BP2 Bbp/53BP2 [Homo sapiens]   | 2e-11   |
| 1169   | Z74972                                | S. cerevisiae chromosome XV reading frame ORF YOR064c  | 2e-11   | 3041855  | (AC004537) similar to tumor suppressor p33ING1; similar to AF044076 (PID:g28929208) [Homo sapiens] | 2e-40   |
| 1170   | Z82181                                | Human DNA sequence from cosmid E86D10 on chromosome 22. contains ESTs, exontrap, complete sequence | 7e-12   | <NONE>   | <NONE>   | <NONE>  |
| 1171   | X77738                                | H.sapiens red cell anion exchanger (EPB3, AE1, Band 3) gene, 3' region                             | 7e-12   | 2135416  | hypothetical protein - human >gi 288145  | 0.012   |
| 1172   | S61977                                | Medium-chain acyl-CoA dehydrogenase {exon 10, intron 10} [human, Genomic, 1407 nt]                 | 6e-12   | 113666   | !!!! ALU CLASS A WARNING ENTRY !!!!  | 0.100   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1173   | X66285                                | M.musculus DNA for HC1 locus  | 6e-12   | 854065   | (X83413) U88 [Human herpesvirus 6]                                  | 2e-06   |
| 1174   | S78744                                | protein S=activated protein C cofactor [rats, liver, mRNA, 3315 nt]       | 6e-12   | 2338292  | (AF009243) proline-rich Gla protein 2 [Homo sapiens]                | 3e-10   |
| 1175   | X58474                                | Bovine OXT gene for oxytocin, 5' noncoding region                         | 2e-12   | 1296429  | (L77967) small proline-rich protein with paired repeat              | 4.1     |
| 1176   | Z56314                                | H.sapiens CpG DNA, clone 10h10, reverse read cpg 10h10.rtl.               | 2e-12   | 2935221  | (AF030154) pVII [bovine adenovirus type 3]                          | 2.8     |
| 1177   | Z56314                                | H.sapiens CpG DNA, clone 10h10, reverse read cpg10h10.rtl.                | 2e-12   | 2708659  | (AF037440) putative 26 kDA protein [Edwardsiella ictaluri]          | 2.8     |
| 1178   | Z19543                                | M.musculus h2-calponin cDNA   | 2e-12   | 2497945  | BETA SCRUIIN >gi 1015535 (Z47541) beta scruiin [Limulus polyphemus] | 2e-04   |
| 1179   | S45332                                | erythropoietin receptor [human, placental, Genomic, 8647 nt]              | 7e-13   | 728835   | !!!! ALU SUBFAMILY SC WARNING ENTRY                                 | 0.074   |
| 1180   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 2e-13   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1181   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 2e-13   | <NONE>   | <NONE>  | <NONE>  |
| 1182   | Z59509                                | H.sapiens CpG DNA, clone 15a1, reverse read cpg15a1.rta.                  | 2e-13   | 3150251  | (AL023634) hypothetical protein   | 0.66    |
| 1183   | D10170                                | Human CYP11B2 gene for steroid 18-hydroxylase                             | 2e-13   | 728837   | !!!! ALU SUBFAMILY SQ WARNING ENTRY   | 3e-05   |
| 1184   | U65416                                | Human MHC class I molecule (MICB) gene, complete cds                      | 2e-13   | 126295   | LINE-1 REVERSE TRANSCRIPTASE HOMOLOG  | 6e-11   |
| 1185   | AJ006031                              | Mus musculus IHABP gene, promoter   | 8e-14   | 2132223  | hypothetical protein YPL186c-yeast  | 1.1     |
| 1186   | U34976                                | Human gamma-sarcoglycan mRNA, complete cds                                | 8e-14   | 1054903  | (U34976) gamma-sarcoglycan [Homo sapiens] >gi 4239660 sapiens]  | 0.034   |
| 1187   | D30647                                | Rat mRNA for very-long-chain Acyl-CoA dehydrogenase, complete cds         | 8e-14   | 3183512  | ACYL-COA DEHYDROGENASE, VERY-LONG-CHAIN SPECIFIC (VLCAD) >gi 2388724 (AF017176) very-long-chain acyl-CoA dehydrogenase [Mus musculus] | 8e-23   |
| 1188   | Z63247                                | H.sapiens CpG DNA, clone 7g4, forward read cpg7g4 fla.                    | 6e-14   | 86285  | Histone H1.01 - chicken   | 6.8     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1189   | U27196                                | Gallus gallus zinc finger protein (Fzf-1) mRNA, complete cds.                  | 3e-14   | 2134436  | zinc finger protein - chicken (fragment)   | 4e-10   |
| 1190   | M26219                                | African green monkey origin of replication                                     | 2e-14   | <NONE>   | <NONE>   | <NONE>  |
| 1191   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                       | 2e-14   | 4235641  | (AF119040) NL0D [Lycopersicon esculentum]  | 0.65    |
| 1192   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds      | 2e-14   | 3043728  | (AB011174) KIAA0602 protein [Homo sapiens]   | 0.28    |
| 1193   | AJ005866                              | Homo sapiens mRNA for putative Sqv-7-like protein, partial                     | 2e-14   | 4008517  | (AJ005866) Sqv-7-like protein [Homo sapiens]                                       | 0.004   |
| 1194   | U32709                                | Haemophilus influenzae Rd section 24 of 163 of the complete genome             | 2e-14   | 3861056  | (AJ235272) POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE (pnp) [Rickettsia prowazekii] | 6e-28   |
| 1195   | AF073485                              | Homo sapiens MHC class I-related protein MR1 precursor (MR1) gene, partial cds | 8e-15   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY   | 1.0     |
| 1196   | AF052135                              | Homo sapiens clone 23625 mRNA sequence   | 8e-15   | 4098124  | (U73522) AMSH [Homo sapiens]   | 8e-14   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1197   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                  | 3e-15   | <NONE>   | <NONE>  | <NONE>  |
| 1198   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 3e-15   | 113671   | !!!! ALU CLASS F WARNING ENTRY !!!!   | 1.7     |
| 1199   | Z75104                                | S.cerevisiae chromosome XV reading frame ORF YOR196c                      | 3e-15   | 3878570  | (Z46381) similar to lipoic acid synthase; cDNA EST yk283b6.3 comes from this gene; cDNA EST yk283b6.5 comes from this gene; cDNA EST yk472f5.3 comes from this gene; cDNA EST yk472f5.5 comes from this gene; cDNA EST yk476e7.3... | 1e-15   |
| 1200   | X70052                                | S.cerevisiae sof1 gene  | 3e-15   | 1125754  | (U42833) coded for by C. elegans cDNA cm16f6; coded for by C. elegans cDNA CEESU63F; similar to S. cerevisiae SOF1 protein (SP:P33750) [Caenorhabditis elegans]   | 3e-29   |
| 1201   | AF012899                              | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 2e-15   | <NONE>   | <NONE>  | <NONE>  |
| 1202   | M92295                                | Gorilla gorilla gamma 1 and gamma-2 globin genes, complete cds.           | 1e-15   | 284078   | hypothetical protein 2 - human >gi 182220   | 7.4     |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|---|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1203                                  | L34587    | Homo sapiens<br>RNA polymerase<br>II elongation<br>factor SIII, p15<br>subunit mRNA,<br>complete cds. ><br>::<br>gb AR022286 A<br>R022 286<br>Sequence 7 from<br>patent US<br>5792634 | 9e-16   | <NONE>  | <NONE>  | <NONE>  |
| 1204                                  | D83649    | Xenopus laevis<br>mRNA for<br>xSox7 protein,<br>complete cds  | 8e-16   | 2447043   | (D83649) xSox7<br>protein [Xenopus<br>laevis]                   | 4e-06   |
| 1205                                  | AC005190  | Homo sapiens<br>PAC clone<br>DJ1152D16<br>from Xq23;<br>complete<br>sequence [Homo<br>sapiens]  | 3e-16   | <NONE>  | <NONE>  | <NONE>  |
| 1206                                  | J03626    | Human UMP<br>synthase mRNA,<br>complete cds.  | 3e-16   | 113667  | !!!! ALU CLASS B<br>WARNING ENTRY<br>!!!!                       | 0.65    |
| 1207                                  | J00083    | Human Alu<br>family<br>interspersed<br>repeat; clone<br>BLUR11.   | 3e-16   | 728836  | !!!! ALU<br>SUBFAMILY SP<br>WARNING ENTRY                       | 4e-06   |
| 1208                                  | U70674    | Mus musculus<br>m-Numb (m-nb)<br>mRNA,<br>complete cds  | 1e-16   | <NONE>  | <NONE>  | <NONE>  |
| 1209                                  | U66619    | Human SWI-<br>SNF complex 60<br>KDa subunit<br>(BAF60c)<br>mRNA,<br>complete cds  | 1e-16   | 1549247   | (U66619) SWI/SNF<br>complex 60 KDa<br>subunit [Homo<br>sapiens] | 0.003   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1210   | U75467                                | Drosophila melanogaster Rga and Atu genes, complete cds                   | 1e-16   | 1658503  | (U75467) Atu [Drosophila melanogaster]   | 5e-32   |
| 1211   | M72709                                | Human alternative splicing factor mRNA, complete cds.                     | 3e-17   | <NONE>   | <NONE>   | <NONE>  |
| 1212   | U26556                                | Human ferritin H (FTHL 13) pseudogene.                                    | 3e-17   | <NONE>   | <NONE>   | <NONE>  |
| 1213   | D32064                                | Human gene for 2-oxoglutarate dehydrogenase, complete cds                 | 3e-17   | 2088843  | (AF003386) F59E12.9 gene product [Caenorhabditis elegans]  | 0.12    |
| 1214   | M76364                                | Human (Papua New Guinean) Mitochondrial DNA control region, sequence 131. | 3e-17   | 114009   | APAG PROTEIN >gi 72927 pir  BVECA G apaG protein - Escherichia coli >gi 40918 (X04711) URF hypothetical protein [Escherichia coli] | 0.006   |
| 1215   | AF017466                              | Homo sapiens genomic sequence from subtelomeric region of chromosome 4q   | 1e-17   | 3947985  | (U78948) MADS-box protein 2 [Malus domestica]  | 4.1     |
| 1216   | AF004876                              | Homo sapiens 54TMp (54tm) mRNA, complete cds                              | 1e-17   | 4101574  | (AF004876) 54TMp [Homo sapiens]  | 0.006   |
| 1217   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                  | 9e-18   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1218   | AF086758                              | Rattus norvegicus Na-K-2C1 cotransporter                          | 4e-18   | 3892703  | (AL033545) putative glycine-rich protein [Arabidopsis thaliana]  | 0.30    |
| 1219   | AF020089                              | Homo sapiens PEN11B mRNA, complete cds                            | 4e-18   | 2642493  | (AF023910) DNA topoisomerase I [Physarum polycephalum]   | 0.083   |
| 1220   | X82333                                | H.sapiens IRLB gene (exon1-3)                                     | 4e-18   | 106837   | ir1B protein - human (fragment) >gi 33969  | 2e-11   |
| 1221   | AB002383                              | Human mRNA for KIAA0385 gene, complete cds                        | 4e-18   | 3228540  | (AF060181) zinc finger protein [Homo sapiens]  | 6e-25   |
| 1222   | X98485                                | P. vivax PV14 gene  | 1e-18   | <NONE>   | <NONE>   | <NONE>  |
| 1223   | Z79057                                | H. sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA21E8   | 1e-18   | 2981631  | (AB012223) ORF2 [Canis familiaris]   | 0.001   |
| 1224   | L01457                                | Homo sapiens (clone JH4B1) PM-scl autoantigen mRNA, complete cds. | 1e-18   | 346287   | nucleolar 100K polymyositis-scleroderma protein - human >gi 35555 (X66113) PM/Scl 100kD nucleolar protein [Homo sapiens] | 0.001   |
| 1225   | L02897                                | Dog nonerythroid beta-spectrin mRNA, 3' end.                      | 4e-19   | 3493358  | (AB017037) nonstructural protein precursor [Himetobi P virus]  | 0.12    |
| 1226   | AB012162                              | Homo sapiens mRNA for APCL protein, complete cds                  | 4e-19   | 3894265  | (AB012162) APCL protein [Homo sapiens]   | 0.002   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1227   | AB011093                              | Homo sapiens mRNA for KIAA0521 protein, partial cds   | 4e-19   | 3043566  | (AB011093) KIAA0521 protein [Homo sapiens]  | 9e-09   |
| 1228   | X78454                                | X.laevis AB21 mRNA for RPD3 homologue   | 4e-19   | 3023945  | HISTONE DEACETYLASE (HD) thaliana]  | 5e-34   |
| 1229   | U88895                                | Human endogenous retrovirus H D1 leader region/integrase-derived ORF1, ORF2, and putative envelope protein mRNA, complete cds | 2e-19   | 59977  | (Z14310) tripartite fusion transcript PLA2L [Human endogenous retrovirus]                     | 1e-04   |
| 1230   | U34377                                | Human tyrosine kinase TXK (txk) gene, exon 13.  | 1e-19   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY  | 3e-05   |
| 1231   | X72966                                | M.musculus rab3A gene   | 1e-19   | 2408076  | (Z99167) putative peroxisomal organisation and biogenesis protein [Schizosaccharomyces pombe] | 2e-09   |
| 1232   | AB007953                              | Homo sapiens mRNA, chromosome 1 specific transcript KIAA0484  | 4e-20   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1233   | D14034                                | Human gene for Zn-alpha2-glycoprotein, complete cds    | 2e-20   | 3928756  | (AB001535) similar to C.elegans hypothetical protein CET01H8.1, CEC05C12.3, CEF54D1.5. similar to trp and trp-like proteins [Homo sapiens]   | 1e-07   |
| 1234   | X82126                                | H.sapiens HOK-2 gene, exon 2                           | 2e-20   | 2137269  | DNA-binding protein - mouse gi 437444  | 1e-19   |
| 1235   | AF093684                              | Luciferase reporter vector pXP2 *SA, complete sequence | 5e-21   | 2773363  | (AF041382) microtubule binding protein D-CLIP-190  | 5.5     |
| 1236   | J05272                                | Human IMP dehydrogenase type 1 mRNA complete cds.      | 5e-21   | 124417   | INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE 1 (IMP DEHYDROGENASE 1) (IMPDH-I) (IMPD 1) I - human  | 2e-04   |
| 1237   | D86997                                | Human (lambda) DNA for immunoglobulin light chain      | 5e-21   | 3878261  | (Z75712) Similarity to S. Pombe BEM1/BUD5 suppressor; cDNA EST EMBL:Z14470 comes from this gene; cDNA EST yk482d4.3 comes from this gene; cDNA EST yk482d4.5 comes from this gene [Caenorhabditis elegans] | 6e-46   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1238   | Z79865                                | H.sapiens chromosome 22 CpG island NDA genomic Msel fragment, clone 302f3, forward read 302f3.f | 2e-21   | 2739037  | (AF024614) ADAM 10 [Caenorhabditis elegans] Zinc-binding metalloprotease domain; cDNA EST CEMSA42F comes from this gene; cDNA EST yk218f3.3 comes from this gene; cDNA EST yk443d9.3 comes from this gene; cDNA EST yk443d9.5 comes from this gene; cDNA...   | 2.6     |
| 1239   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 6e-22   | 3924779  | (AL008585) similar to laminin B, cDNA EST yk450d8.5 comes from this gene; cDNA EST yk249a6.5 comes from this gene; cDNA EST yk219a2.5 comes from this gene; cDNA EST yk355e4.5 comes from this gene; cDNA EST yk224f4.5 comes fr... >gi 3924881 gnl PID e 1354569 from this gene; cDNA EST yk249a6.5 comes from this gene; cDNA EST yk219a2.5 comes from this gene; cDNA EST yk355e4.5 comes from this gene; cDNA EST yk224f4.5 comes from... | 0.35    |
| 1240   | U67824                                | Human primary Alu transcript  | 6e-22   | 728832   | !!!! ALU SUBFAMILY SB WARNING ENTRY   | 5e-07   |
| 1241   | AF070636                              | Homo sapiens clone 24686 mRNA sequence  | 2e-22   | 98710  | Fatty-acid synthase (EC 2.3.1.85) - Brevibacterium ammoniagenes   | 2.5     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1242   | D14034                                | Human gene for Zn-alpha2-glycoprotein, complete cds                                   | 2e-22   | 4185939  | (Y17832) pol protein [Human endogenous retrovirus K]   | 0.29    |
| 1243   | M61835                                | Human lactase phlorizin hydrolase (LCT) gene, exon 2.                                 | 2e-22   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY   | 0.006   |
| 1244   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 6e-23   | 1350828  | RABPHILIN-3A >gi 477100 pir  A48097 rabphilin-3A - bovine >gi 285646 gnl PID d1003285                                  | 0.14    |
| 1245   | AF074985                              | Homo sapiens full length insert cDNA YH73H06  | 8e-24   | 3170548  | (AF056116) unknown [Fugu rubripes]   | 0.24    |
| 1246   | D14878                                | Human mRNA for protein D123, complete cds   | 7e-24   | <NONE>   | <NONE>   | <NONE>  |
| 1247   | D16917                                | Human HepG2 3' region cDNA, clone hmd3d07   | 6e-24   | 1397345  | (U61955) contains multiple region of strong similarity to C2H2-type zinc fingers (PS:PS00028) [Caenorhabditis elegans] | 2.4     |
| 1248   | Z69654                                | Human DNA sequence from cosmid L98A6, Huntington's Disease Region, chromosome 4p16.3. | 3e-24   | 4240566  | (AF123462) neurexin III [Homo sapiens]   | 4.5     |
| 1249   | AB007914                              | Homo sapiens mRNA for KIAA0445 protein, complete cds                                  | 2e-24   | 3885949  | (AF095568) amelogenin [Paleosuchus palpebrosus]  | 3.2     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1250   | AF088072                              | Homo sapiens full length insert cDNA clone ZD93D10   | 2e-24   | 323091   | Immunodominant microneme protein Etp100 - Eimeria tenella >gi 2707733 (AF032905) microneme protein precursor Etmic-1 [Eimeria tenella] | 0.34    |
| 1251   | AF069489                              | Homo sapiens cAMP specific phosphodiesterase 4A variant pde46 (PDE4A) gene, exons 2 through 13 and alternative splice exons 3a, 6a, 6b, and 9a | 2e-24   | 728836   | !!!! ALU SUBFAMILY SP WARNING ENTRY  | 1e-05   |
| 1252   | Y12853                                | Homo sapiens P2X7 gene, exon 4-8   | 9e-25   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY   | 1e-05   |
| 1253   | M27830                                | Human 28S ribosomal RNA gene, complete cds.  | 8e-25   | <NONE>   | <NONE>   | <NONE>  |
| 1254   | AB007953                              | Homo sapiens mRNA, chromosome 1 specific transcript KIAA0484   | 8e-25   | <NONE>   | <NONE>   | <NONE>  |
| 1255   | Z60212                                | H.sapiens CpG DNA, clone 195c8, forward read cpg 195c8.ftla.   | 8e-25   | 158154   | (M81959) POU domain protein [Drosophila melanogaster]  | 3.3     |
| 1256   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds   | 7e-25   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1257   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 7e-25   | <NONE>   | <NONE>   | <NONE>  |
| 1258   | Y12851                                | Homo sapiens P2X7 gene, exon 1 and joined CDS                                   | 2e-25   | <NONE>   | <NONE>   | <NONE>  |
| 1259   | U64033                                | Mus musculus Tera (Tera) mRNA, complete cds                                     | 9e-26   | <NONE>   | <NONE>   | <NONE>  |
| 1260   | U19181                                | Rattus norvegicus Rabin3 mRNA, complete cds.                                    | 9e-26   | 624225   | (U19181) Rabin3 [Rattus norvegicus]  | 1e-13   |
| 1261   | AF020788                              | Caenorhabditis elegans SEL-10 (sel-10) mRNA, complete cds                       | 9e-26   | 3915881  | SEL-10 PROTEIN<br>Candida CDC4 gene (TR:E234056); cDNA EST EMBL:D27699 comes from this gene; cDNA EST EMBL:D27698 comes from this gene; cDNA EST EMBL:D32793 comes from this gene; cDNA EST EMBL:D33271 comes from this gen... | 7e-32   |
| 1262   | AB016930                              | Cricetulus griseus mRNA for Phosphatidylglycerophosphate synthase, complete cds | 8e-26   | 4159682  | (AB016930) Phosphatidylglycerophosphate synthase [Cricetulus griseus]  | 0.045   |
| 1263   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds  | 3e-26   | 3878629  | (Z93385) predicted using Genefinder; Similarity to B.subtilis GTP-binding protein  | 2e-10   |
| 1264   | X91195                                | H.sapiens SOM172 mRNA   | 1e-26   | <NONE>   | <NONE>   | <NONE>  |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1265   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                       | 1e-26   | 1360637  | (X95995) ENBP1 [Vicia sativa]   | 3.1     |
| 1266   | L08237                                | Human MG21 mRNA, partial cds.  | 1e-26   | 950411   | (L08237) located at OATL1 [Homo sapiens]  | 9e-09   |
| 1267   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                       | 9e-27   | 3881080  | (AL032657) similar to EGF-like domain; cDNA EST yk299a12.3 comes from this gene; cDNA EST EMBL:D35398 comes from this gene; cDNA EST yk331h6.5 comes from this gene; cDNA EST yk299a12.5 comes from this gene; cDNA EST yk467g8.... | 0.001   |
| 1268   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                       | 8e-27   | 1731324  | HYPOTHETICAL PROTEIN >gi 166306   | 4.0     |
| 1269   | X89211                                | H.sapiens DNA for endogenous retroviral like element                           | 8e-27   | 2065209  | (Y12713) Gag polyprotein [Mus musculus]   | 0.005   |
| 1270   | U73166                                | Homo sapiens cosmid clone LUCA15 from 3p21.3, complete sequence [Homo sapiens] | 3e-27   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY  | 4e-04   |
| 1271   | D78255                                | Mouse mRNA for PAP-1, complete cds   | 3e-27   | 1850098  | (D78255) PAP-1 [Mus musculus]   | 2e-10   |
| 1272   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                                       | 1e-27   | 2133579  | spermatophorin Sp23 - yellow mealworm molitor]  | 0.39    |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1273                                  | AB015202  | Homo sapiens gene for hippocalcin, exon 2, 3 and complete cds      | 1e-27   | 3877698   | (Z83318) predicted using Genefinder; cDNA EST yk369e7.5 comes from this gene [Caenorhabditis elegans]   | 0.37    |
| 1274                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds                           | 1e-27   | 3328188   | (AF074902) laminin alpha chain [Caenorhabditis elegans]   | 0.19    |
| 1275                                  | Z29336    | H.sapiens gene for Cu/Zn-superoxide dismutase                      | 1e-27   | 728831  | !!!! ALU SUBFAMILY J WARNING ENTRY  | 6e-05   |
| 1276                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds                           | 9e-28   | 2133579   | spermatophorin Sp23 - yellow mealworm molitor]  | 9.2     |
| 1277                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds                           | 9e-28   | 2133579   | spermatophorin Sp23 - yellow mealworm molitor]  | 0.054   |
| 1278                                  | AB001636  | Homo sapiens mRNA for ATP-dependent RNA helicase #46, complete cds | 4e-28   | 3913425   | PUTATIVE PRE-MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE >gi 2275203 (AC002337) RNA helicase isolog [Arabidopsis thaliana]  | 3e-22   |
| 1279                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds                           | 3e-28   | 4056454   | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 0.066   |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|--|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION                              | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1280                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 3e-28   | 4056454   | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 4e-05   |
| 1281                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>  | <NONE>  | <NONE>  |
| 1282                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>  | <NONE>  | <NONE>  |
| 1283                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>  | <NONE>  | <NONE>  |
| 1284                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>  | <NONE>  | <NONE>  |
| 1285                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>  | <NONE>  | <NONE>  |
| 1286                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | <NONE>  | <NONE>  | <NONE>  |
| 1287                                  | AF100694  | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 140505  | PROBABLE INTRON MATURASE liverwort (Marchantia polymorpha) chloroplast >gi 11663  | 3.0     |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1288   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                  | 1e-28   | 140505   | PROBABLE INTRON MATURASE liverwort (Marchantia polymorpha) chloroplast >gi 11663  | 1.8     |
| 1289   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                  | 1e-28   | 2133579  | spermatophorin Sp23 - yellow mealworm molitor]  | 0.50    |
| 1290   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                  | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 0.087   |
| 1291   | Z63029                                | H.sapiens CpG DNA, clone 77b3, forward read cpg77b3.ftla. | 1e-28   | 2493240  | HYPOTHETICAL 29.3 KD PROTEIN pseudotsugata nuclear polyhedrosis virus]  | 0.014   |
| 1292   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                  | 1e-28   | 118588   | DEHYDRIN DHN3 >gi 100035 pir  S1813 9 dehydrin DHN3 - garden pea >gi 20709 (X63063) pea dehydrin DHN3 [Pisum sativum]   | 0.010   |
| 1293   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                  | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 0.007   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION                              | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1294   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 0.002   |
| 1295   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 126363   | LAMININ ALPHA-1 CHAIN PRECURSOR precursor - human   | 3e-04   |
| 1296   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 1e-04   |
| 1297   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 3e-05   |
| 1298   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 3157926  | (AC002131) Strong similarity to extensin-like protein gb Z34465 from Zea mays. [Arabidopsis thaliana]   | 2e-05   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION                              | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1299   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 1e-05   |
| 1300   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 320919   | kinetoplast-associated protein - Trypanosoma cruzi >gi 162142 (M25364) kinetoplast-associated protein   | 1e-07   |
| 1301   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 9e-08   |
| 1302   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 1e-09   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION                              | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1303   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 9e-10   |
| 1304   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 4e-10   |
| 1305   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 9e-11   |
| 1306   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 1e-28   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana] | 6e-11   |
| 1307   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds | 4e-29   | <NONE>   | <NONE>  | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1308   | AF079529                              | Homo sapiens cAMP-specific phosphodiesterase 8B            | 4e-29   | <NONE>   | <NONE>  | <NONE>  |
| 1309   | X93334                                | H.sapiens mitochondrial DNA, complete genome               | 4e-29   | 116977   | CYTOCHROME C OXIDASE POLYPEPTIDE I chain I - human mitochondrion (SGC1) >gi 13006 (V00662) cytochrome oxidase I [Homo sapiens] >gi 506829 (J01415) cytochrome oxidase subunit I [Homo sapiens] sapiens] | 3e-09   |
| 1310   | AF020760                              | Homo sapiens serine protease (Omi) mRNA, complete cds      | 4e-29   | 2738915  | (AF020760) serine protease [Homo sapiens]   | 8e-12   |
| 1311   | U95097                                | Xenopus laevis mitotic phosphoprotein 43 mRNA, partial cds | 4e-29   | 2072294  | (U95097) mitotic phosphoprotein 43 [Xenopus laevis]   | 1e-25   |
| 1312   | L32162                                | Homo sapiens transcription factor mRNA, 5' end.            | 2e-29   | 2501706  | RENAL TRANSCRIPTION FACTOR KID-1 finger protein [Mus musculus]  | 8e-15   |
| 1313   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds                   | 1e-29   | 4056454  | (AC005990) Contains repeated region with similarity to gb U43627 extensin (atExt1) gene from Arabidopsis thaliana. ESTs gb Z34165 and gb Z18788 come from this gene. [Arabidopsis thaliana]             | 1e-04   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1314   | AF100694                              | Mus musculus Pontin52 mRNA, complete cds           | 1e-29   | 1169643  | FMRFAMIDE-RELATED NEUROPEPTIDES PRECURSOR >gi 416208 (U03137) neuropeptide precursor FMRFamide-related peptide [Lymnaea stagnalis]                   | 1e-05   |
| 1315   | U50839                                | Homo sapiens g16 protein (g16) mRNA, complete cds  | 1e-29   | 3212101  | (AF069517) RNA binding protein DEF-3 [Homo sapiens]  | 6e-10   |
| 1316   | X69711                                | H. sapiens mRNA for ICAM-R                         | 5e-30   | 299356   | Intercellular adhesion molecule 3, ICAM-3=lymphocyte function-associated antigen 1 counter-receptor homolog [human, tonsil, Peptide Partial, 518 aa] | 3e-08   |
| 1317   | AF010227                              | Homo sapiens receptor-associated coactivator 3     | 5e-30   | 2331250  | (AF012108) Amplified in Breast Cancer [Homo sapiens]   | 8e-09   |
| 1318   | AF086395                              | Homo sapiens full length insert cDNA clone ZD75C01 | 2e-30   | 3861241  | (AJ235273) CELL SURFACE ANTIGEN (sca5)   | 4.2     |
| 1319   | M27830                                | Human 28S ribosomal RNA gene, complete cds.        | 2e-30   | 1730522  | PHOSPHOGLYCERATE KINASE 2.7.2.3) - Pyrococcus woesei >gi 1054832 (X73527) phosphoglycerate kinase [Pyrococcus woesei]                                | 3.8     |
| 1320   | M79307                                | Mouse GTP-binding protein (Rab17) mRNA sequence.   | 2e-30   | 464564   | RAS-RELATED PROTEIN RAB-17 Rab17 - mouse (fragment) >gi 297157 (X70804) rab17 [Mus musculus]   | 9e-11   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1321   | AL022168                              | Human DNA sequence from clone U247E12 on chromosome Xq22-23, complete sequence [Homo sapiens] | 1e-30   | 2072967  | (U93570) putative p150 [Homo sapiens]   | 3e-11   |
| 1322   | X85124                                | M.musculus pacsin gene  | 1e-30   | 2217964  | (Z50798) p52 [Gallus gallus]  | 1e-34   |
| 1323   | U37408                                | Homo sapiens phosphoprotein CtBP mRNA, complete cds   | 5e-31   | 74518  | structural polypeptide - Venezuelan equine encephalitis virus (strain TRD) >gi 323710 (J04332) poly-envelope protein [Venezuelan equine encephalitis virus] | 1.1     |
| 1324   | L04193                                | Human lens membrane protein (mp19) gene, exon 11.   | 2e-31   | 728831   | !!!! ALU SUBFAMILY J WARNING ENTRY  | 7e-07   |
| 1325   | M11167                                | Human 28S ribosomal RNA gene.   | 6e-32   | <NONE>   | <NONE>  | <NONE>  |
| 1326   | M33336                                | Human cAMP-dependent protein kinase type I-alpha subunit (PRKAR1A) mRNA, complete cds         | 2e-32   | <NONE>   | <NONE>  | <NONE>  |
| 1327   | J03060                                | Human glucocerebrosidase pseudogene, complete cds   | 2e-32   | 2144479  | glucosylceramidase (EC 3.2.1.45) precursor - human  | 1e-05   |
| 1328   | U33053                                | Human lipid-activated protein kinase PRK1 mRNA, complete cds                                  | 7e-33   | 2137689  | protein kinase (EC 2.7.1.37) - mouse  | 1e-14   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1329   | J04617                                | Human elongation factor EF-1-alpha gene, complete cds. > :: dbj E02629 E02629 DNA of human polypeptide chain elongation factor-1 alpha | 6e-33   | <NONE>   | <NONE>   | <NONE>  |
| 1330   | L40396                                | Homo sapiens (clone s22i71) mRNA fragment  | 6e-33   | 124235   | INTERMEDIATE FILAMENT PROTEIN B protein B - common roundworm                               | 1.00    |
| 1331   | Z72813                                | S.cerevisiae chromosome VII reading frame ORF YGR028w  | 6e-33   | 1709135  | MSP1 PROTEIN HOMOLOG Yeast MSP1 protein (TAT-binding homolog 4)                            | 8e-50   |
| 1332   | AB007941                              | Homo sapiens mRNA for KIAA0472 protein, partial cds  | 2e-33   | 1150834  | (U42471) Wiscott-Aldrich Syndrome protein homolog [Mus musculus]                           | 2.0     |
| 1333   | AF044574                              | Rattus norvegicus putative peroxisomal 2,4-dienoyl-CoA reductase (DCR-AKL) mRNA, complete cds  | 2e-34   | 4105269  | (AF044574) putative peroxisomal 2,4-dienoyl-CoA reductase [Rattus norvegicus]              | 6e-15   |
| 1334   | D14657                                | Human mRNA for KIAA0101 gene, complete cds   | 7e-35   | <NONE>   | <NONE>   | <NONE>  |
| 1335   | X69910                                | H.sapiens p63 mRNA for transmembrane protein   | 7e-35   | 2136323  | trithorax homolog HTX - human (fragment) homolog=MLL {alternative splicing, clone 14p-18B} | 0.94    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1336   | AF053455                              | Homo sapiens tetraspan TM4SF (TSPAN-5) gene, complete cds                | 7e-35   | 3152703  | (AF065389) tetraspan NET-4 [Homo sapiens]           | 1e-25   |
| 1337   | X58374                                | D.melanogaster crn mRNA  | 3e-35   | 117478   | CROOKED NECK PROTEIN                                | 6e-41   |
| 1338   | AF086492                              | Homo sapiens full length insert cDNA clone ZD95D11                       | 9e-36   | 2909809  | (AF031328) aminoglycoside 6'-N-acetyltransferase It | 1.9     |
| 1339   | Z96223                                | H.sapiens telomeric DNA sequence, clone 12PTEL120, read 12PTELOO120. seq | 3e-36   | 2408068  | (Z99165) hypothetical protein                       | 0.61    |
| 1340   | Z37986                                | H.sapiens mRNA for phenylalkylamine binding protein.                     | 1e-36   | 1362793  | emopamil-binding protein - human >gi 780263         | 5e-11   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1341   | U57847                                | Human ribosomal protein S27 mRNA, complete cds. end similar to similar to metallopanstimulin 1 > :: gb AA316327 AA316327 EST188061 HCC cell line (matatasis to liver in mouse) II Homo sapiens cDNA 5' end similar to similar to metallopanstimulin 1 | 3e-37   | 1171014  | 40S RIBOSOMAL PROTEIN S27 growth factor-inducible zinc finger protein MPS-1 - human >gi 431319 (L19739) metallopanstimulin [Homo sapiens] >gi 1373421 (U57847) ribosomal protein S27 | 1.4     |
| 1342   | Y15054                                | Rattus norvegicus mRNA for 70 kDa tumor specific antigen, partial   | 3e-37   | 3123027  | 70 KD WD-REPEAT TUMOR-SPECIFIC ANTIGEN >gi 2505957 gnl PID e353992 (Y15054) 70 kD tumor-specific antigen [Rattus norvegicus]   | 2e-15   |
| 1343   | AF084205                              | Rattus norvegicus serine/threonine protein kinase TAO1 mRNA, complete cds   | 3e-37   | 3452473  | (AF084205) serine/threonine protein kinase TAO1 [Rattus norvegicus]  | 5e-47   |
| 1344   | X78604                                | R.norvegicus (Sprague Dawley) ARL5 mRNA for ARF-like protein 5  | 1e-37   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1345   | AJ236644                              | Homo sapiens chromosome 22 CpG island DNA, genomic Mse I fragment, clone 22CGIB49A3, complete read                           | 1e-37   | 2239219  | (Z97210) hypothetical protein  | 6e-05   |
| 1346   | U09367                                | Human zinc finger protein ZNF136   | 4e-39   | 2137269  | DNA-binding protein - mouse >gi 437444                                       | 7e-23   |
| 1347   | Z69649                                | Human DNA sequence from cosmid L69F7B, Huntington's Disease Region, chromosome 4p16.3 contains Huntington Disease (HD) gene. | 3e-39   | 3096918  | (AL023094) putative cyclase associated protein CAP [Arabidopsis thaliana]    | 5.6     |
| 1348   | AF065389                              | Homo sapiens tetraspan NET-4 mRNA, complete cds  | 1e-39   | 3152703  | (AF065389) tetraspan NET-4 [Homo sapiens]                                    | 6e-29   |
| 1349   | AF038172                              | Homo sapiens clone 23923 mRNA sequence   | 1e-40   | 1813464  | (U60883) CapC [Bacillus firmus]  | 2.8     |
| 1350   | Z83095                                | H.sapiens Fanconi anaemia group A gene, exons 39, 40, 41, 42 and 43  | 1e-40   | 2137870  | zinc finger protein - mouse (fragment)                                       | 3e-23   |
| 1351   | AF057734                              | Homo sapiens 17-beta-hydroxysteroid dehydrogenase IV (HSD17B4) gene, exon 16   | 1e-40   | 2842416  | (AL008730) dJ487J7.1.1 (putative protein dJ487J7.1 isoform 1) [Homo sapiens] | 6e-61   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1352   | AF070567                              | Homo sapiens clone 24544 beta-dystrobrevin mRNA, partial cds                   | 4e-41   | 3133087  | (Y15718) dystrobrevin B DTN-B2 [Homo sapiens]  | 7e-13   |
| 1353   | AF006088                              | Homo sapiens Arp2/3 protein complex subunit p16-Arc (ARC16) mRNA, complete cds | 2e-41   | 3121767  | ARP2/3 COMPLEX 16 KD SUBUNIT   | 3e-36   |
| 1354   | X69942                                | M.musculus mRNA of enhancer-trap-locus 1                                       | 6e-42   | 2291152  | (AF016418) No definition line found [Caenorhabditis elegans]   | 6.4     |
| 1355   | X87838                                | H.sapiens mRNA for beta-catenin  | 5e-42   | 1373019  | (U28811) cysteine-rich fibroblast growth factor receptor   | 8e-05   |
| 1356   | AB018268                              | Homo sapiens mRNA for KIAA0725 protein, partial cds                            | 5e-42   | 3882171  | (AB018268) KIAA0725 protein [Homo sapiens]   | 2e-33   |
| 1357   | M84424                                | Human cathepsin E (CTSE) gene, exon 9 and complete cds.                        | 2e-42   | <NONE>   | <NONE>   | <NONE>  |
| 1358   | U80776                                | Human EST clone NIB1543 mariner transposon Hsmar1 orf gene, complete cds       | 2e-42   | 2231380  | (U80776) orf; encodes putative chimeric protein with SET domain in N-terminus with similarity to several other human, Drosophila, nematode and yeast proteins [Homo sapiens] | 3e-11   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1359   | U55184                                | Human G protein Golf alpha gene, exon 12 and complete cds                    | 2e-42   | 3165531  | (AF067608) No definition line found [Caenorhabditis elegans]  | 1e-16   |
| 1360   | AC005190                              | Homo sapiens PAC clone DJ1152D16 from Xq23, complete sequence [Homo sapiens] | 6e-43   | 2978255  | (AB007407) myeloid zinc finger protein-2 [Mus musculus]   | 2.3     |
| 1361   | AB018284                              | Homo sapiens mRNA for KIAA0741 protein, complete cds                         | 5e-43   | <NONE>   | <NONE>  | <NONE>  |
| 1362   | AB011137                              | Homo sapiens mRNA for KIAA0565 protein, complete cds                         | 5e-43   | 3043654  | (AB011137) KIAA0565 protein [Homo sapiens]  | 1e-07   |
| 1363   | M93651                                | Human set gene, complete cds.  | 2e-43   | <NONE>   | <NONE>  | <NONE>  |
| 1364   | Z47087                                | H.sapiens mRNA for RNA polymerase II elongation factor-like protein.         | 2e-43   | 1872514  | (U84404) E6-associated protein E6-AP/ubiquitin-protein ligase [Homo sapiens] >gi 2361031 (AF016708) E6-AP ubiquitin-protein ligase [Homo sapiens] | 7.2     |
| 1365   | U27197                                | Drosophila melanogaster pelota (pelo) mRNA, complete cds                     | 2e-43   | 1352736  | PELOTA PROTEIN >gi 973224 (U27197) pelota [Drosophila melanogaster]   | 1e-46   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1366   | D80007                                | Human mRNA for KIAA0185 gene, partial cds                                   | 6e-44   | 2498864  | RRP5 PROTEIN HOMOLOG (KIAA0185) hypothetical protein YM9959.11C of <i>S.cerevisiae</i> . [Homo sapiens]   | 6e-09   |
| 1367   | AF005039                              | Homo sapiens secretory carrier membrane protein (SCAMP3) mRNA, complete cds | 6e-44   | 2232243  | (AF005039) secretory carrier membrane protein [Homo sapiens]  | 2e-09   |
| 1368   | X68101                                | <i>R.norvegicus</i> trg mRNA  | 2e-44   | 550420   | (X68101) trg gene product [ <i>Rattus norvegicus</i> ]  | 1e-37   |
| 1369   | AF044206                              | Homo sapiens cyclooxygenase (COX-2) gene, promoter and exon 1               | 2e-45   | 2072953  | (U93565) putative p150 [Homo sapiens]   | 5e-06   |
| 1370   | L48708                                | Homo sapiens facio-genital dysplasia (FGD1) gene, 5' end of intron 17       | 8e-46   | <NONE>   | <NONE>  | <NONE>  |
| 1371   | X15822                                | Human COX VIIa-L mRNA for liver-specific cytochrome c oxidase (EC 1.9.3.1.) | 3e-46   | 117121   | CYTOCHROME C OXIDASE POLYPEPTIDE VIIA-LIVER PRECURSOR >gi 2144370 pir  OSH U7L cytochrome-c oxidase (EC 1.9.3.1) chain VIIa precursor, hepatic - human >gi 30147 (X15822) precursor (AA -23 to 60) [Homo sapiens] | 5e-13   |
| 1372   | U47323                                | Mus musculus stromal cell protein mRNA, complete cds                        | 3e-46   | 1493833  | (U47323) stromal cell protein [Mus musculus]  | 1e-48   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1373   | AF059524                              | Homo sapiens reticulon gene family protein                                     | 7e-47   | 1731169  | HYPOTHETICAL 113.1 KD PROTEIN T28D9.7 IN CHROMOSOME II >gi 861264 (U28738) coded for by C. elegans cDNA yk8h5.3; coded for by C. elegans cDNA yk8h5.5; similar to C. elegans deg-1 and mec-4 in exon 2 [Caenorhabditis elegans] | 7.8     |
| 1374   | AJ132583                              | Homo sapiens mRNA for puromycin sensitive aminopeptidase, partial              | 3e-47   | 1777519  | (U39123) T cell receptor beta chain [Homo sapiens]  | 9.7     |
| 1375   | M97856                                | Homo sapiens histone binding protein mRNA, complete cds.                       | 3e-47   | 2645327  | (U83821) NADH dehydrogenase subunit 3 [Oryzomys palustris]  | 5.7     |
| 1376   | U53220                                | Human retinoblastoma-related Rb2/p130 gene, 5' flanking region and partial cds | 3e-47   | 2499225  | CMP-SIALIC ACID TRANSPORTER CMP-sialic acid transporter [Cricetulus griseus]  | 5.3     |
| 1377   | X87870                                | H. sapiens mRNA for hepatocyte nuclear factor 4a                               | 1e-47   | 728832   | !!!! ALU SUBFAMILY SB WARNING ENTRY   | 7.3     |
| 1378   | AF060195                              | Mus musculus proteasome regulator PA28 beta subunit gene, complete cds         | 3e-48   | 478681   | limb deformity protein - chicken  | 0.25    |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1379   | AB018285                              | Homo sapiens mRNA for KIAA0742 protein, partial cds   | 1e-48   | 3122969  | TESTIS SPECIFIC PROTEIN A (ZINC FINGER PROTEIN TSGA)<br>>gi 281040 pir  S28499 probable zinc finger protein - rat >gi 57504(X59993) zinc finger protein | 1e-30   |
| 1380   | U35032                                | Human endogenous retrovirus clone c5.11, HERV-H multiply spliced subgenomic leader, protease and integrase region mRNA, partial cds | 4e-49   | 88558  | retroviral proteinase-like protein - human  | 6e-05   |
| 1381   | AB007956                              | Homo sapiens mRNA, chromosome 1 specific transcript KIAA0487  | 1e-49   | <NONE>   | <NONE>  | <NONE>  |
| 1382   | D86987                                | Homo sapiens mRNA for KIAA0214 protein, complete cds  | 1e-49   | 2497944  | ALPHA SCRUIIN<br>>gi 633238 (Z38132) scruiin [Limulus polyphemus]<br>>gi 1093326 prf  2103269A scrulin [Limulus sp.]                                    | 9.7     |
| 1383   | U25826                                | Human transcription factor (SC1) gene, complete cds.  | 4e-50   | <NONE>   | <NONE>  | <NONE>  |
| 1384   | U46690                                | Mus musculus ATP-dependent RNA helicase mRNA, partial cds.  | 4e-50   | 1335873  | (U46690) ATP-dependent RNA helicase [Mus musculus]  | 3e-24   |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1385                                  | AF072128  | Mus musculus claudin-2 mRNA, complete cds   | 2e-50   | 3335184   | (AF072128) claudin-2 [Mus musculus]                                      | 4e-24   |
| 1386                                  | AF093593  | Homo sapiens snRNA activating protein complex 19kDa subunit (SNAP19) mRNA, complete cds | 1e-50   | 3668416   | (AF093593) snRNA activating protein complex 19kDa subunit [Homo sapiens] | 0.003   |
| 1387                                  | U79745    | Homo sapiens monocarboxylate transporter homologue MCT6 mRNA, complete cds              | 1e-50   | 1177607   | (X92485) pval [Plasmodium vivax]   | 2e-07   |
| 1388                                  | L09647    | Rattus norvegicus hepatocyte nuclear factor 3a  | 1e-50   | 404764  | (L10409) fork head related protein [Mus musculus]                        | 2e-21   |
| 1389                                  | X61506    | Mouse E46 mRNA for E46 protein  | 4e-51   | 114909  | BRAIN PROTEIN E46  | 1e-20   |
| 1390                                  | M33387    | Human debrisoquine 4-hydroxylase (CYP2D8P) and  | 1e-51   | 126296  | LINE-1 REVERSE TRANSCRIPTASE HOMOLOG protein [Nycticebus coucang]        | 5e-15   |
| 1391                                  | AF019767  | Homo sapiens zinc finger protein (ZPR1) mRNA, complete cds                              | 4e-52   | 961507  | (D63788) anchor protein, LCM   | 5.9     |
| 1392                                  | Z37986    | H.sapiens mRNA for phenylalkylamine binding protein.                                    | 2e-52   | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1393   | U65416                                | Human MHC class I molecule (MICB) gene, complete cds                         | 2e-52   | 3878637  | (Z49128) weak similarity with SINR protein (Swiss Prot accession number P06533); cDNA EST EMBL:T00631 comes from this gene; cDNA EST yk293d10.5 comes from this gene [Caenorhabditis elegans] | 8.7     |
| 1394   | Z57647                                | H.sapiens CpG DNA, clone 189a6, forward read cpg189a6.ftla.                  | 2e-52   | 111187   | beta-globin DNA-binding protein B1, transcription factor PU.1 - mouse >gi 200586 (M32370) PU.1 protein [Mus musculus >gi 200972 (M38252) transcription factor Pu.1 [Mus musculus]             | 5.8     |
| 1395   | L13738                                | Human activated p21cdc42Hs kinase (ack) mRNA, complete cds.                  | 2e-52   | 2921447  | (AF037260) non-receptor protein tyrosine kinase Ack [Mus musculus]  | 7e-23   |
| 1396   | AF042379                              | Homo sapiens spindle pole body protein spc97 homolog GCP2 mRNA, complete cds | 7e-53   | 2801701  | (AF042379) spindle pole body protein spc97 homolog GCP2   | 1e-16   |
| 1397   | AF047441                              | Homo sapiens RNA polymerase I 40kD subunit mRNA, complete cds                | 6e-53   | 3914807  | DNA-DIRECTED RNA POLYMERASE I 40 KD POLYPEPTIDE (RPA40) (RPA39) >gi 2266929 (AF008442) RNA polymerase I subunit hRPA39 [Homo sapiens]   | 4e-19   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1398   | AF104670                              | Homo sapiens cell cycle protein (PA2G4) gene, exons 6 through 13, and complete cds   | 2e-53   | <NONE>   | <NONE>   | <NONE>  |
| 1399   | S60754                                | {VNTR locus DXZ4, hypervariable tandem repeat cluster} [human, Genomic, 2991 nt] > :: gb L07935 HUM VNTRA Homo sapiens microsatellite VNTR DNA sequence. | 2e-53   | 1209669  | (U38810) CAGR1 [Homo sapiens] >gi 3098420 (AF040945) homeotic regulator homolog MAB21 [Mus musculus] | 4.6     |
| 1400   | D86972                                | Human mRNA for KIAA0218 gene, complete cds   | 1e-53   | 3426041  | (AC005168) unknown protein [Arabidopsis thaliana]  | 9.1     |
| 1401   | AJ236682                              | Homo sapiens chromosome 22 CpG island DNA, genomic Mse 1 fragment, clone 22CGIB49E6, complete read   | 7e-54   | 3928721  | (AL034355) putative cytochrome oxidase subunit I [Streptomyces coelicolor]                           | 0.30    |
| 1402   | AJ236682                              | Homo sapiens chromosome 22 CpG island DNA, genomic Mse1 fragment, clone 22CGIB49E6, complete read  | 6e-54   | 3928721  | (AL034355) putative cytochrome oxidase subunit I [Streptomyces coelicolor]                           | 0.28    |
| 1403   | M37583                                | Human histone (H2A.Z) mRNA, complete cds.  | 6e-54   | 70711  | histone H2A.F, embryonic - chicken   | 2e-16   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1404   | AJ009947                              | Homo sapiens mRNA for putative ATPase, partial                               | 6e-54   | 3550295  | (AJ009947) putative ATPase [Homo sapiens]               | 3e-18   |
| 1405   | Y08459                                | B.taurus mRNA for novel cytoplasmic protein                                  | 2e-54   | <NONE>   | <NONE>  | <NONE>  |
| 1406   | AF042384                              | Homo sapiens BC-2 protein mRNA, complete cds                                 | 2e-54   | 2828147  | (AF042384) BC-2 protein [Homo sapiens]                  | 2e-14   |
| 1407   | AF042379                              | Homo sapiens spindle pole body protein spc97 homolog GCP2 mRNA, complete cds | 8e-55   | 2801701  | (AF042379) spindle pole body protein spc97 homolog GCP2 | 2e-17   |
| 1408   | AF005355                              | Oryctolagus cuniculus translation initiation factor eIF2C mRNA, complete cds | 7e-55   | 3253159  | (AF005355) translation initiation factor eIF2C          | 3e-53   |
| 1409   | AF008442                              | Homo sapiens RNA polymerase I subunit hRPA39 mRNA, complete cds              | 3e-55   | 3335138  | (AF047441) RNA polymerase I 40kD subunit [Homo sapiens] | 3e-20   |
| 1410   | AF047441                              | Homo sapiens RNA polymerase I 40kD subunit mRNA, complete cds                | 3e-55   | 3335138  | (AF047441) RNA polymerase I 40kD subunit [Homo sapiens] | 3e-20   |
| 1411   | X08004                                | Human mRNA for Rap1B protein > :: emb A08693 A08693 H.sapiens rap1b cDNA     | 2e-55   | 539995   | transforming protein rap1b - rat (strain Copenhagen)    | 2e-18   |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|---|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1412                                  | AF010403  | Homo sapiens ALR mRNA, complete cds                 | 2e-55   | 2358285   | (AF010403) ALR [Homo sapiens]  | 1e-49   |
| 1413                                  | M77016    | Human tropomodulin mRNA, complete cds.              | 8e-56   | 262249  | (S52010) orf1 5' of EpoR [mice, Peptide, 85 aa] [Mus sp.]                | 0.027   |
| 1414                                  | AB020633  | Homo sapiens mRNA for KIAA0826 protein, partial cds | 2e-56   | <NONE>  | <NONE>   | <NONE>  |
| 1415                                  | X87489    | H.sapiens genomic DNA (chromosome 3; clone NL1243D) | 2e-56   | 1814029   | (U84501) cuticle collagen [Caenorhabditis briggsae]                      | 0.038   |
| 1416                                  | AB007893  | Homo sapiens KIAA0433 mRNA, partial cds             | 2e-56   | 2887437   | (AB007893) KIAA0433 [Homo sapiens]                                       | 9e-21   |
| 1417                                  | X78925    | H.sapiens HZF2 mRNA for zinc finger protein         | 1e-56   | 3342002   | (AF054180) hematopoietic cell derived zinc finger protein [Homo sapiens] | 2e-21   |
| 1418                                  | Z56281    | H.sapiens mRNA for interferon regulatory factor 3   | 9e-57   | 2497442   | INTERFERON REGULATORY FACTOR 3 factor 3 [Homo sapiens]                   | 2e-21   |
| 1419                                  | U78772    | Homo sapiens nuclear VCP-like protein NVLp.1        | 8e-57   | 2406565   | (U68140) nuclear VCP-like protein NVLp.2 [Homo sapiens]                  | 5e-20   |
| 1420                                  | D79994    | Human mRNA for KIAA0172 gene, partial cds           | 3e-57   | 1136404   | (D79994) similar to ankyrin of Chromatium vinosum. [Homo sapiens]        | 9e-38   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|---|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1421   | AB002342                              | Human mRNA for KIAA0344 gene, complete cds                                      | 1e-57   | 2224629  | (AB002342) KIAA0344 [Homo sapiens]   | 4e-20   |
| 1422   | L19437                                | Human transaldolase mRNA containing transposable element, complete cds          | 1e-57   | 1553119  | (U63159) transaldolase [Mus musculus]  | 2e-20   |
| 1423   | D17532                                | Human mRNA for RCK, complete cds  | 9e-58   | 129376   | PROBABLE ATP-DEPENDENT RNA HELICASE P54 (ONCOGENE RCK) (DEAD BOX PROTEIN 6)                                      | 1e-10   |
| 1424   | X79568                                | H.sapiens BDP1 mRNA for protein-tyrosine-phosphatase                            | 9e-58   | 1871531  | (X79568) protein-tyrosine-phosphatase  | 1e-22   |
| 1425   | X79568                                | H.sapiens BDP1 mRNA for protein-tyrosine-phosphatase                            | 9e-58   | 1871531  | (X79568) protein-tyrosine-phosphatase  | 9e-23   |
| 1426   | AB012295                              | Homo sapiens HKE1.5 mRNA for GDS-related protein, complete cds                  | 7e-58   | 2648021  | (Z97184) RGL2 [Homo sapiens]   | 9e-19   |
| 1427   | AF086040                              | Homo sapiens full length insert cDNA clone YX52E07                              | 1e-58   | 543222   | glutamine (Q)-rich factor 1, QRF-1 - mouse factor 1, QRF-1 [mice, B-cell leukemia, BCL1, Peptide Partial, 84 aa] | 3e-36   |
| 1428   | AB018195                              | Homo sapiens ca xi mRNA for carbonic anhydrase-related protein XI, complete cds | 4e-59   | <NONE>   | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1429   | AF071777                              | Mus musculus IRE1 (Ire1) mRNA, complete cds  | 4e-59   | 3766209  | (AF071777) IRE1 [Mus musculus]   | 7e-28   |
| 1430   | AB000462                              | Homo sapiens mRNA for SH3 binding protein, complete cds, clone:RES4-23A  | 3e-59   | <NONE>   | <NONE>   | <NONE>  |
| 1431   | AF038172                              | Homo sapiens clone 23923 mRNA sequence   | 3e-59   | 3758855  | (Z98551)MAL3P6.11 [Plasmodium falciparum]  | 1.3     |
| 1432   | Z84812                                | Human DNA sequence from phage pTEL from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3 Contains ESTs | 1e-59   | 400927   | RIBONUCLEOPROTEIN RB97D ribonucleoprotein [Drosophila melanogaster]  | 2.5     |
| 1433   | U36484                                | Human laminin-binding protein gene, partial cds, and E2 small nucleolar RNA gene, complete sequence                                    | 1e-59   | 226005   | protein 40kD [Mus musculus]  | 7e-05   |
| 1434   | L11285                                | Homo sapiens ERK activator kinase (MEK2) mRNA.   | 1e-59   | 2499630  | DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE KINASE 2 (MAP KINASE KINASE 2) (MAPKK 2) kinase type 2 [Gallus gallus] | 3e-21   |
| 1435   | AF086555                              | Homo sapiens full length insert cDNA clone ZE14E04   | 4e-60   | 3287674  | (AC005239) F23149_1 [Homo sapiens]   | 2e-04   |

| Nearest Neighbor (BlastN vs. Genbank) |           |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|---------------------------------------|-----------|--|---------|---|--|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1436                                  | M24766    | Human (clone pHAIV2-12) alpha-2 collagen type IV                           | 4e-60   | 29551   | (X05610) alpha (2) chain [Homo sapiens]  | 6e-15   |
| 1437                                  | X65550    | H.sapiens mki67a mRNA (long type) for antigen of monoclonal antibody Ki-67 | 4e-60   | 1170654   | ANTIGEN KI-67 >gi 539555 pir  A4866 6 cell proliferation antigen Ki-67, long form - human Ki-67 [Homo sapiens] | 3e-15   |
| 1438                                  | M27319    | Human calmodulin mRNA, complete cds.                                       | 4e-60   | 1345451   | (X05949) Calmodulin (AA 2 - 59) (449 is 1st base in codon) [Drosophila melanogaster]                           | 7e-20   |
| 1439                                  | Y12781    | Homo sapiens mRNA for transducin (beta) like 1 protein                     | 3e-60   | 62133   | (X06172) put. 134 kD protein (AA 1 - 1187); put. replicase   | 7.4     |
| 1440                                  | AB002383  | Human mRNA for KIAA0385 gene, complete cds                                 | 1e-60   | 1001548   | (D64000) hypothetical protein  | 4.4     |
| 1441                                  | AF070614  | Homo sapiens clone 24732 unknown mRNA, partial cds                         | 2e-61   | 3283879   | (AF070614) unknown [Homo sapiens]  | 3e-17   |
| 1442                                  | AB002326  | Human mRNA for KIAA0328 gene, partial cds                                  | 6e-62   | 547891  | MICROTUBULE-ASSOCIATED PROTEIN 4 microtubule-associated protein-U [Bos taurus]                                 | 5.6     |
| 1443                                  | AF086471  | Homo sapiens full length insert cDNA clone ZD88A01                         | 5e-62   | <NONE>  | <NONE>   | <NONE>  |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1444   | AB002311                              | Human mRNA for KIAA0313 gene, complete cds                                      | 2e-62   | 2506357  | 2,3-DIHYDROXYPHENYLPROPIONATE 1,2-DIOXYGENASE<br>>gi 1657544 (U73857)<br>similar to mcpI gene (catechol 2,3-dioxygenase) of A. eutrophus 3-(2,3-dihydroxyphenylpropionate) 1, 2-dioxygenase 2,3-dihydroxyphenylpropionate 1,2-dioxygenase | 3.4     |
| 1445   | AF069737                              | Xenopus laevis notchless (nle) mRNA, complete cds                               | 2e-62   | 3687833  | (AF069737) notchless [Xenopus laevis]   | 1e-55   |
| 1446   | AF044209                              | Homo sapiens nuclear receptor co-repressor N-CoR mRNA, complete cds             | 5e-63   | 2137603  | nuclear receptor co-repressor N-CoR - mouse musculus<br>>gi 1583865 prf 2121436A thyroid hormone receptor co-repressor [Mus musculus]   | 2e-47   |
| 1447   | M69238                                | Human aryl hydrocarbon receptor nuclear translocator (ARNT) mRNA, complete cds. | 2e-63   | 2702319  | (AF001307) aryl hydrocarbon receptor nuclear translocator; Arnt [Homo sapiens]  | 5e-19   |
| 1448   | X80497                                | H.sapiens PHKLA mRNA  | 2e-63   | 1170685  | PHOSPHORYLASE B KINASE ALPHA REGULATORY CHAIN, LIVER ISOFORM (PHOSPHORYLASE KINASE ALPHA L SUBUNIT)<br>>gi 663010 (X80497)<br>phosphorylase kinase phosphorylase kinase alpha subunit [Homo sapiens]                                      | 5e-22   |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|---|---------|---|---|---------|
| SEQ ID                                | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1449                                  | AF031141  | Homo sapiens ubiquitin conjugating enzyme   | 2e-63   | 2623260   | (AF031141) ubiquitin conjugating enzyme [Homo sapiens]  | 1e-23   |
| 1450                                  | Z37166    | H.sapiens BAT1 mRNA for nuclear RNA helicase  | 6e-64   | 2500529   | PROBABLE ATP-DEPENDENT RNA HELICASE P47 >gi 2135840 pir  I37201 nuclear RNA helicase (DEAD family) BAT1 - human >gi 587146 (Z37166) nuclear RNA helicase (DEAD family) [Homo sapiens] | 9e-24   |
| 1451                                  | M64240    | Human helix-loop-helix zipper protein (max) mRNA, complete cds. >:: gb I41138 I41138 Sequence 1 from patent US 5624818 >:: gb I77062 I77062 Sequence 1 from patent US 5693487 | 5e-64   | 88175   | Myc-binding factor Max, short form - human  | 8e-22   |
| 1452                                  | M98252    | Homo sapiens lysyl hydroxylase (partial clone 2.2 Kb LH) RNA, complete mature peptide.  | 2e-64   | 400205  | PROCOLLAGEN-LYSINE,2-OXOGLUTARATE 5-DIOXYGENASE PRECURSOR (LYSYL HYDROXYLASE) lysyl hydroxylase [Homo sapiens]  | 7e-22   |
| 1453                                  | U09550    | Human oviductal glycoprotein mRNA, complete cds.  | 8e-65   | 2493676   | OVIDUCT-SPECIFIC GLYCOPROTEIN PRECURSOR (OVIDUCTAL GLYCOPROTEIN) (OVIDUCTIN)  | 2e-11   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1454   | X67877                                | R.norvegicus mRNA for cytosolic resiniferatoxin-binding protein                            | 7e-65   | 423664   | resiniferatoxin-binding protein RBP-26, cytosolic - rat >gi 311660 (X67877) cytosolic resiniferatoxin binding protein RBP-26 [Rattus norvegicus >gi 1093373 prf 2103310A resiniferatoxin-binding protein [Rattus norvegicus] | 2e-40   |
| 1455   | AB018254                              | Homo sapiens mRNA for KIAA0711 protein, complete cds                                       | 6e-65   | 92298  | glutamine/glutamic acid-rich protein   | 0.98    |
| 1456   | J03607                                | Human 40-kDa keratin intermediate filament precursor gene.                                 | 3e-65   | 1070608  | keratin 19, type I, cytoskeletal - human sapiens]  | 4e-07   |
| 1457   | U65896                                | Human gamma-glutamyl carboxylase gene, complete cds  | 2e-65   | <NONE>   | <NONE>   | <NONE>  |
| 1458   | U07681                                | Human NAD(H)-specific isocitrate dehydrogenase alpha subunit precursor mRNA, complete cds. | 2e-65   | 1708399  | ISOCITRATE DEHYDROGENASE (NAD), MITOCHONDRIAL SUBUNIT ALPHA PRECURSOR (ISOCITRIC DEHYDROGENASE) (NAD+-SPECIFIC ICDH) dehydrogenase alpha chain precursor - human >gi 706839 subunit precursor [Homo sapiens]                 | 4e-26   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1459   | U88080                                | Human zinc finger protein (LD5-1) gene, exons 4, 5 and 6, and complete cds | 2e-65   | 1373394  | (U57796) zinc finger protein [Homo sapiens] >gi 2306773  | 2e-39   |
| 1460   | M96625                                | Gallus domesticus tensin mRNA sequence.                                    | 3e-66   | 2134419  | tensin - chicken (fragment) >gi 63805 (Z18529) tensin [Gallus gallus] >gi 212755 (L06662) tensin [Gallus gallus] | 1e-51   |
| 1461   | U13262                                | Mus musculus myelin gene expression factor (MEF-2) mRNA, partial cds.      | 1e-70   | 536926   | (U13262) myelin gene expression factor [Mus musculus]  | 9e-42   |
| 1462   | U64033                                | Mus musculus Tera (Tera) mRNA, complete cds                                | 5e-72   | 1575505  | (U64033) Tera [Mus musculus]   | 9e-34   |
| 1463   | X78989                                | M.musculus mRNA for testin   | 6e-74   | 1351218  | TESTIN 2 (TES2) [CONTAINS: TESTIN 1  | 8e-31   |
| 1464   | U64033                                | Mus musculus Tera (Tera) mRNA, complete cds                                | 2e-74   | 1575505  | (U64033) Tera [Mus musculus]   | 5e-37   |
| 1465   | AF057365                              | Canis familiaris UDPN-acetylglucosamine transporter mRNA, complete cds     | 9e-79   | 3298605  | (AF057365) UDP N-acetylglucosamine transporter [Canis familiaris]  | 9e-10   |
| 1466   | AJ006064                              | Rattus norvegicus mRNA for coronin-like protein                            | 1e-82   | 3757680  | (AJ006064) coronin-like protein [Rattus norvegicus]  | 3e-62   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1467   | U91582                                | Macaca fascicularis UDP-glucuronosyltransferase mRNA, complete cds                       | 4e-89   | 140396   | KARYOGAMY PROTEIN KAR4 yeast (Saccharomyces cerevisiae)  | 1e-08   |
| 1468   | X06762                                | Mouse Hox2.3 mRNA  | 3e-92   | 123255   | HOMEBOX PROTEIN HOX-B7 (HOX-2C)  | 9e-23   |
| 1469   | AB016930                              | Cricetulus griseus mRNA for Phosphatidylglycerophosphate synthase, complete cds          | 5e-94   | 4159682  | (AB016930) Phosphatidylglycerophosphate synthase [Cricetulus griseus]  | 7e-34   |
| 1470   | X74504                                | M.musculus T10 mRNA  | 7e-97   | 1711658  | SER/THR-RICH PROTEIN T10 IN DGCR REGION<br>>gi 480900 pir  S3748 8 gene T10 protein - mouse  | 3e-59   |
| 1471   | U13175                                | Rattus norvegicus clone ubc10a ubiquitin conjugating enzyme (E217kB) mRNA, complete cds. | 3e-98   | 1351345  | UBIQUITIN-CONJUGATING ENZYME E2-17 KD 3 (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN) (E2(17)KB 3)<br>>gi 1085588 pir  S533 58 ubiquitin conjugating enzyme (E217kB) - rat<br>>gi 595666 (U13175) ubiquitin conjugating enzyme [Rattus norvegicus] norvegicus<br>>gi 1145691 (U39318) UbCH5C [Homo sapiens] | 5e-05   |



| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|--|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1472   | S79873                                | h-lamp-2=lysosome-associated membrane protein-2 protein-2b (LAMP2) mRNA, alternatively spliced form h-lamp-2b, complete cds. | e-119   | <NONE>   | <NONE>  | <NONE>  |
| 1473   | D13623                                | Rat mRNA for p34 protein, complete cds   | e-112   | 480379   | ribosome-binding protein p34-rat sp.]   | 2e-05   |
| 1474   | AB013357                              | Mus musculus mRNA for 49 kDa zinc finger protein, complete cds   | e-136   | 4153886  | (AB013357) 49 kDa zinc finger protein   | 5e-08   |
| 1475   | AB016930                              | Cricetulus griseus mRNA for Phosphatidylglycerophosphate synthase, complete cds  | e-117   | 4159682  | (AB016930) Phosphatidylglycerophosphate synthase [Cricetulus griseus]   | 4e-32   |
| 1476   | U38253                                | Rattus norvegicus initiation factor eIF-2B gamma subunit (eIF-2B gamma) mRNA, complete cds                                   | e-103   | 2494312  | TRANSLATION INITIATION FACTOR EIF-2B GAMMA SUBUNIT (EIF-2B GDP-GTP EXCHANGE FACTOR) subunit [Rattus norvegicus] | 3e-42   |
| 1477   | X73683                                | R.norvegicus mRNA for histone H3.3   | e-117   | 122075   |   | 1e-45   |
| 1478   | U32498                                | Rattus norvegicus rsec8 mRNA, partial cds  | e-108   | 2143962  | rsec8 - rat (fragment) >gi 1019441 (U32498) rsec8 [Rattus norvegicus]   | 7e-48   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1479   | U41736                                | Mus musculus ancient ubiquitous 46 kDa protein AUP1 precursor (Aup1) mRNA, complete cds | e-146   | 1517822  | (U41736) ancient ubiquitous 46 kDa protein AUP46 precursor [Mus musculus] | 5e-49   |
| 1480   | AF041338                              | Bos taurus vacuolar proton pump subunit SFD alpha isoform (SFD) mRNA, complete cds      | e-119   | 2895578  | (AF041338) vacuolar proton pump subunit SFD alpha isoform [Bos taurus]    | 3e-49   |
| 1481   | AF064553                              | Mus musculus NSD1 protein mRNA, complete cds  | e-121   | 3329465  | (AF064553) NSD1 protein [Mus musculus]                                    | 2e-50   |
| 1482   | AB000517                              | Rattus sp. mRNA for CDP-diacylglycerol synthase, complete cds                           | e-146   | 1517822  | (U41736) ancient ubiquitous 46 kDa protein AUP46 precursor [Mus musculus] | 2e-51   |
| 1483   | D38517                                | Mouse mRNA for Dhml protein, complete cds   | e-118   | 2137562  | mouse Dhml protein - mouse musculus]                                      | 6e-54   |
| 1484   | X54352                                | M.domesticus MD6 mRNA   | e-139   | 1085499  | CDC4 repeat unit-containing protein - mouse                               | 1e-55   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|-----------|---------------------------------------|---|---------|---|---|---------|
|           | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1485      | U57692                                | Mus musculus<br>N-terminal<br>asparagine<br>amidohydrolase<br>(Ntan1) mRNA,<br>complete cds | e-118   | 2498797   | PROTEIN N-<br>TERMINAL<br>ASPARAGINE<br>AMIDOHYDROLAS<br>E (PROTEIN NH2-<br>TERMINAL<br>ASPARAGINE<br>DEAMIDASE) (NTN-<br>AMIDASE) (PNAD)<br>(PROTEIN NH2-<br>TERMINAL<br>ASPARAGINE<br>AMIDOHYDROLAS<br>E) (PNAA)<br>>gi 1373365 (U57691)<br>N-terminal asparagine<br>amidohydrolase [Mus<br>musculus]<br>amidohydrolase [Mus<br>musculus] | 5e-57   |
| 1486      | X80169                                | M.musculus<br>mRNA for 200<br>kD protein  | e-119   | 1717793   | PROTEIN TSG24<br>(MEIOTIC CHECK<br>POINT<br>REGULATOR)<br>>gi 1083553 pir  A551<br>17 tsg24   | 9e-58   |

| SEQ<br>ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |  |         |
|-----------|---------------------------------------|--|---------|---|--|---------|
|           | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION   | DESCRIPTION  | P VALUE |
| 1487      | U57692                                | Mus musculus<br>N-terminal<br>asparagine<br>amidohydrolase<br>(Ntan1) mRNA,<br>complete cds            | e-120   | 2498797   | PROTEIN N-<br>TERMINAL<br>ASPARAGINE<br>AMIDOHYDROLAS<br>E (PROTEIN NH2-<br>TERMINAL<br>ASPARAGINE<br>DEAMIDASE) (NTN-<br>AMIDASE) (PNAD)<br>(PROTEIN NH2-<br>TERMINAL<br>ASPARAGINE<br>AMIDOHYDROLAS<br>E) (PNA A)<br>>gi 1373365 (U57691)<br>N-terminal asparagine<br>amidohydrolase [Mus<br>musculus]<br>amidohydrolase [Mus<br>musculus] | 8e-58   |
| 1488      | U08215                                | Mus musculus<br>Hsp70-related<br>NST-1 (hsr.1)<br>mRNA,<br>complete cds.                               | e-109   | 473407  | (U08215) NST-1 [Mus<br>musculus]   | 7e-58   |
| 1489      | D85926                                | Mouse mRNA<br>for Ray,<br>complete cds   | e-110   | 1944389   | (D85926) Ray [Mus<br>musculus]   | 2e-58   |
| 1490      | L20427                                | Rattus<br>norvegicus<br>dihydroxypolypr<br>enylbenzoate<br>methyltransferas<br>e mRNA,<br>complete cds | e-123   | 457372  | (L20427)<br>dihydroxypolyprenylb<br>enzoate<br>methyltransferase<br>dihydroxypolyprenylb<br>enzoate<br>methyltransferase<br>[Rattus norvegicus]  | 4e-59   |
| 1491      | X56044                                | M.musculus<br>mRNA for<br>protein Htf9C  | e-121   | 3183977   | (X56044) protein<br>Htf9C [Mus musculus]   | 1e-60   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |  |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |  |         |
|--------|---------------------------------------|--|---------|--|--|---------|
|        | ACCESSION                             | DESCRIPTION  | P VALUE | ACCESSION  | DESCRIPTION  | P VALUE |
| 1492   | S74774                                | p59fyn(T)=OKT 3-induced calcium influx regulator                               | e-163   | 729896   | PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (P59-FYN)<br>>gi 420217 pir  A44991 protein-tyrosine kinase (EC 2.7.1.112)<br>fyn - mouse             | 8e-63   |
| 1493   | U88873                                | Mus musculus BUB2-like protein 1 (HBLP1) mRNA, complete cds                    | e-123   | 4099611  | (U88873) BUB2-like protein 1 [Mus musculus]  | 1e-63   |
| 1494   | U48852                                | Cricetulus griseus HT protein mRNA, complete cds.                              | e-117   | 1216486  | (U48852) HT protein [Cricetulus griseus]   | 7e-64   |
| 1495   | AF032667                              | Rattus norvegicus rexo70 mRNA, complete cds                                    | e-142   | 2827160  | (AF032667) rexo70 [Rattus norvegicus]  | 5e-66   |
| 1496   | M62722                                | Chinese hamster phosphatidylserine decarboxylase mRNA, 3' end.                 | e-114   | 118910   | PHOSPHATIDYLSERINE DECARBOXYLASE PROENZYME<br>>gi 109423 pir  A38732 phosphatidylserine decarboxylase (EC 4.1.1.65) - Chinese hamster (fragment) | 2e-67   |
| 1497   | AF072758                              | Mus musculus fatty acid transport protein 3 mRNA, partial cds                  | e-130   | 3335567  | (AF072758) fatty acid transport protein 3; FATP3 [Mus musculus]  | 1e-67   |
| 1498   | AB005549                              | Rattus norvegicus mRNA for atypical PKC specific binding protein, complete cds | e-113   | 3868778  | (AB005549) atypical PKC specific binding protein [Rattus norvegicus]   | 2e-69   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1499   | U57344                                | Mus musculus homeobox protein Meis3 mRNA, complete cds                          | e-143   | 3024124  | HOMEBOX PROTEIN MEIS3   | 6e-72   |
| 1500   | U09874                                | Mus musculus SKD3 mRNA, complete cds.   | e-142   | 2493735  | SKD3 PROTEIN SKD3 [Mus musculus]  | 1e-72   |
| 1501   | U72194                                | Mus musculus muskelin mRNA, complete cds  | e-148   | 3493462  | (U72194) muskelin [Mus musculus]  | 2e-74   |
| 1502   | X80169                                | M.musculus mRNA for 200 kD protein  | e-155   | 1717793  | PROTEIN TSG24 (MEIOTIC CHECK POINT REGULATOR) >gi 1083553 pir  A55117 tsg24 | 3e-77   |
| 1503   | U72194                                | Mus musculus muskelin mRNA, complete cds  | e-154   | 3493462  | (U72194) muskelin [Mus musculus]  | 2e-78   |
| 1504   | Y12836                                | Cricetulus griseus mRNA for Zn finger factor                                    | e-146   | 3150148  | (Y12836) Zn finger factor [Cricetulus griseus]                              | 3e-83   |
| 1837   | X80169                                | M.musculus mRNA for 200 kD protein  | e-122   | 1717793  | PROTEIN TSG24 (MEIOTIC CHECK POINT REGULATOR) >gi 1083553 pir  A55117 tsg24 | 2e-56   |
| 1838   | AF080568                              | Rattus norvegicus CTP:phosphoethanolamine cytidyltransferase mRNA, complete cds | e-119   | 3396102  | (AF080568) CTP:phosphoethanolamine cytidyltransferase                       | 6e-58   |

| SEQ ID | Nearest Neighbor (BlastN vs. Genbank) |   |         | Nearest Neighbor (BlastX vs. Non-Redundant Proteins) |   |         |
|--------|---------------------------------------|---|---------|--|---|---------|
|        | ACCESSION                             | DESCRIPTION   | P VALUE | ACCESSION  | DESCRIPTION   | P VALUE |
| 1839   | X99145                                | C.familiaris mRNA for C3VS protein  | e-121   | 1429314  | (X99145) overexpressed in thyroid tissue after TSH stimulation [Canis familiaris]         | 2e-58   |
| 1840   | AF019075                              | Pan troglodytes breast and ovarian cancer susceptibility (BRCA1) gene, partial cds                              | e-145   | 2218154  | (AF005068) breast and ovarian cancer susceptibility protein splice variant [Homo sapiens] | 1e-58   |
| 1841   | U55042                                | Bos taurus myosin X, complete cds   | e-122   | 1755049  | (U55042) myosin X [Bos taurus]  | 1e-61   |
| 1842   | AJ007780                              | Mus musculus mRNA for poly(ADP-ribose polymerase-2  | e-119   | 3283975  | (AF072521) poly-(ADPribosyl)-transferase homolog PARP                                     | 4e-62   |
| 1843   | AF072865                              | Rattus norvegicus thioredoxin reductase (TrxR2) mRNA, nuclear gene encoding mitochondrial protein, complete cds | e-105   | 3757888  | (AF072865) thioredoxin reductase [Rattus norvegicus]                                      | 3e-62   |
| 1844   | U55042                                | Bos taurus myosin X, complete cds   | e-121   | 1755049  | (U55042) myosin X [Bos taurus]  | 1e-62   |
| 1845   | X61506                                | Mouse E46 mRNA for E46 protein  | e-139   | 114909   | BRAIN PROTEIN E46   | 9e-67   |

| Nearest Neighbor (BlastN vs. Genbank) |           |   |         | Nearest Neighbor<br>(BlastX vs. Non-Redundant Proteins) |   |         |
|---------------------------------------|-----------|---|---------|---|---|---------|
| SEQ<br>ID                             | ACCESSION | DESCRIPTION   | P VALUE | ACCESSION   | DESCRIPTION   | P VALUE |
| 1846                                  | D90335    | Bovine mRNA<br>for GTP-binding<br>protein alpha-<br>subunit | e-148   | 585174  | GUANINE<br>NUCLEOTIDE-<br>BINDING PROTEIN,<br>ALPHA-14 SUBUNIT<br>(GL1)<br>>gi 108711 pir  A4089<br>1 GTP-binding protein<br>GL1 alpha chain-<br>bovine protein, alpha-<br>subunit [Bos taurus] | 2e-69   |
| 1847                                  | U49507    | Mus musculus<br>B6CBA Lisch7<br>mRNA, partial<br>cds.       | e-140   | 2121326   | (AC002128) Lisch7<br>[Homo sapiens]   | 2e-74   |